

How can big data analytics help understand migrant integration?

Tuba Bircan,¹ Albert Ali Salah² and Alina Sirbû³
April 2023⁴

1 Introduction

Migration has always been an integral part of history and will certainly shape the future of societies. Migrant integration and inclusion are complex phenomena influenced by multiple and interlinked factors, and they present both challenges and opportunities for global sustainable development. Recent migration trends highlight the significance of recognizing and tackling questions regarding diversity and social cohesion across different world regions. Successful inclusion of immigrants is a prerequisite for social cohesion and economic progress.

This is becoming more pronounced at the national level as the focus on migration intensifies, but it is also evident at the international level. The Global Compact for Safe, Orderly and Regular Migration includes among its 23 objectives the need to collect and utilize accurate data for evidence-informed policymaking. The 2030 Agenda for Sustainable Development explicitly recognizes the contribution of migration to sustainable development. Implementing the Agenda's core principle to "leave no one behind" requires data disaggregation by migratory status, highlighting significant needs for migration data, as well as a "comprehensive strategy for improving migration data at local, national, regional and global levels". Countries are expected to improve their capacity to generate timely, reliable and comparable data on migration to help guide policymakers in

¹ Tuba Bircan is with the Department of Sociology of Vrije Universiteit Brussel, Belgium. She is the corresponding author and can be contacted at tuba.bircan@vub.be.

² Albert Ali Salah is with the Department of Information and Computing Sciences of Utrecht University, the Netherlands. He can be contacted at a.a.salah@uu.nl. He is also affiliated with the Department of Computer Engineering of Boğaziçi University, Turkey.

³ Alina Sirbû is with the Department of Computer Science of University of Pisa, Italy. She can be contacted at alina.sirbu@unipi.it.

⁴ This is the uncorrected author proof. Please cite as "Bircan, T., A.A. Salah, A. Sirbu, 'How Can Big Data Analytics Help Understanding Migrant Integration?' in M. Rango (ed), Practitioners' Guide on Harnessing Data Innovation for Migration Policy, In Press (2023)

devising evidence-based policies and plans of action to tackle migration aspects of the sustainable development goals (SDGs). Data sources typically used in studies of diversity and social cohesion include censuses, administrative data, and surveys including a range of questions about attitudes towards and experiences of immigration and integration, as well as other topics relating to community life. Good practices based on traditional sources and methods for studying migrant integration are evident.

Adequate data are key for evidence-based policymaking. However, while a large amount of official statistics is produced across European Union member States, only a small part of the complexity of migrant integration phenomena can be captured through such data. Social security databases to assess labour market integration allow counting migrants only after they officially enter the systems. On the other hand, indicators to explain the different aspects of migrants' living conditions and monitor integration policies implemented by countries cannot be captured through register data; large-scale surveys such as the European Union Labour Force Survey and the European Union Survey on Income and Living Conditions are the major data sources for these aims. However, migration and migrants have never been the centre of attention (with the exception of some occasional ad hoc modules) for cross-sectional/multinational surveys. These are designed to cover the general resident population in a given country, and although the migrants/non-nationals are incorporated in samples, the representation and coverage of those migrants have been questionable. Consequently, to improve understanding of migrant integration, it is necessary to go beyond traditional data sources, and novel approaches are required.

The use of new data sources for understanding diversity and social cohesion is very recent. Developing social indicators through the use of big data and artificial intelligence technologies is worth exploring, as big data can complement official statistics and surveys, in particular by filling some of the gaps in traditional data sources and methods.

Understanding the various dimensions of the governmental and public commitment to migrant integration is essential to design targeted interventions and, ultimately, to enable migrants and refugees to fully engage with society. For this purpose, useful indicators should be derived from large data repositories. Several past efforts used data science and machine-learning techniques to study migrant integration via different indicators. To give more specific examples, global bilateral remittance flows were mapped using big data visualization techniques for the World Bank's 2011 bilateral remittance database, which estimates flows across 215 countries. Another experimental approach was implemented by PeaceTech Lab, where social media data from South Sudan were analysed and machine-learning-based visualizations for Myanmar were used to better monitor hate speech and public sentiment (Monroe, 2018). An Italian study analysed retail data from a supermarket chain and observed immigrants' food consumption baskets to understand to what extent immigrants converge with or diverge from the norms and habits of the destination country (Sirbu et al., 2021). Other examples include, but are not limited to, the analysis of localized language use and diversity in Twitter data using half a billion geotagged tweets (Magdy et al., 2014), and the analysis of a data set combining mobile phone data

with media events data and data on housing market transactions in Turkey to determine refugees settlement locations (Bansak et al., 2018). This chapter will contribute to the existing efforts by elaborating two special cases, namely the use of social media data and mobile phone data for studying various aspects of migrant and refugee integration.

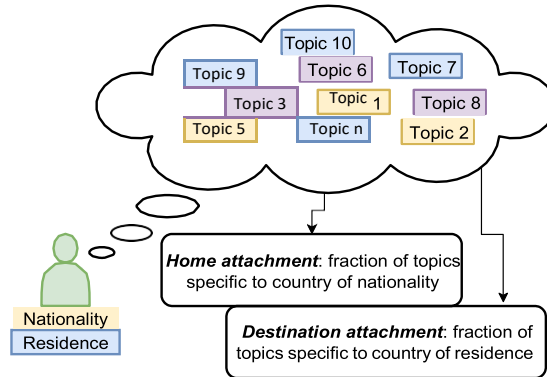
2 Case studies

2.1 Measuring acculturation through Twitter data analysis

The Twitter platform hosts millions of users exchanging daily information in the form of short messages. Besides the messages themselves, the platform also contains the social network of the users. Much of this information is public and can be downloaded using the public Twitter application programming interfaces (APIs), resulting in data that can be employed to study migrant acculturation. Acculturation can be defined as “the process of group and individual changes in culture and behaviour that result from intercultural contact” (Berry, 2019), which is always present in the case of human migration. Acculturation does not merely imply that a culture is absorbed by another, but it is more of a continuous and bilateral process where cultures evolve through contact. We can thus observe different acculturation patterns in different individuals, and Twitter can be employed to study these, as outlined below.

A first step in studying migrant behaviour, including the process of acculturation, is to identify migrants in the data set. While in traditional migration data, this step is intrinsic to the data-collection process, where participants of censuses or surveys are asked to declare their nationality and/or country of birth, along with country of residence, big data from social media and other human digital traces often do not include this information. This is also the case with Twitter, where no information on residence or nationality is available. This means that a preliminary analysis needs to be performed to assign these labels. This is typically done based on location information but also based on the social network, using the location of social contacts (Kim et al., 2020). With the nationality and residence labels assigned, migrants can be defined as users who have a country of residence different from their country of nationality (or birth).

Figure 1. Computing home attachment and destination attachment for Twitter users



Note: The users are assigned a nationality and residence. Then the topics they discuss are studied. Topics specific to the country of residence and nationality are used to compute the home attachment and destination attachment indices.

Once migrant users are selected, acculturation can be measured in various ways. One possible approach is studying the topics that they discuss. If migrants are observed to be involved in conversations specific to their country of origin, they are assumed to maintain a cultural link to their home country. Participation in conversations related to their country of residence, on the other hand, is a sign of absorbing the culture of the destination country. We can identify topics of conversation through the hashtags that users attach to their messages. To each hashtag, we can assign a nationality based on which country has native-born users who post that hashtag most frequently.

The study of the topics discussed provides information on multiple aspects related to the process of acculturation: how much information about a specific country reaches a certain individual; how much an individual is interacting with peers, i.e. social integration; how much an individual is engaged politically and civically, etc. Technically, two indices can be defined: home attachment and destination attachment (Kim et al., 2021), as displayed in Figure 1. Home attachment is the fraction of topics of a user that are specific to their country of nationality. Similarly, destination attachment is the fraction of topics of a user that are specific to their country of residence. Identification of topics on Twitter is easily performed by extracting hashtags, while the country of the topic can be assigned by looking at the nationality of the non-migrant users tweeting with those hashtags.

Figure 2. Conceptualizing the four types of acculturation using home attachment (HA) and destination attachment (DA)

	Low DA	High DA
Low HA	Marginalization	Assimilation
High HA	Separation	Integration

The concepts of home attachment and destination attachment help identify the type of acculturation process that migrants are undergoing (Berry, 2019) by considering specific regions in the space defined by the two dimensions, as outlined in Figure 2. *Separation* is assumed when destination attachment is low and home attachment is high for those persons who maintain strong links with their home country and do not appear to build links to their destination. At the opposite side, we have *assimilation*, with high destination attachment and low home attachment (i.e. those individuals who are immersed completely in the destination culture and lose contact with their origins). Two other known acculturation types are possible. *Integration* applies to high destination attachment and high home attachment (i.e. those individuals who maintain links to their home country while also building links with their destination). *Marginalization* appears when both destination attachment and home attachment are low, meaning that individuals lose connection to their home country and do not build new connections with their destination. The separation among the various types of acculturation can be done by defining strict thresholds on the two indices; however, we believe that they are more useful if considered as a continuum, as a means to place a person on a spectrum of acculturation rather than assigning a unique label.

The method cited here for studying acculturation has an important advantage in that it can be applied globally, in all countries where Twitter is used, and with space and time resolutions that can be arbitrarily low or high. This can complement analyses based on traditional data where resolution and coverage are limited by the complexity of the data-collection process. Additionally, various aspects of acculturation can be analysed separately. While all topics are discussed here, one could also concentrate only on political or other socially relevant topics to zoom in on very specific dimensions of acculturation.

The application of this method does not come without challenges or disadvantages. While Twitter data are public, downloading large amounts of data requires time, specific computer science skills and infrastructure. Data collection and analysis are further complicated by frequent changes to the Twitter API. Furthermore, the API offers only a subsample of the total data, which is further reduced by the fact that geolocation, required to assign residence and nationality of users, is present only in a minority of tweets. In addition, our final analysis is based only on users whose country of residence is different from their country of nationality. As such, while initial Twitter data sets can be very large, multiple stages of filtration can result in rather small final data sets. Hence a lot of effort needs to be allocated to the data-collection phase, which should not be overlooked when planning an analysis.

Another challenge relates to privacy concerns. While using census or survey data to study migration is an established and regulated procedure, the Twitter platform is relatively new, and there are no established frameworks for using data from Twitter or other social media platforms for analysis of public policy issues. The European General Data Protection Regulation includes very strict rules for user profiling used to make automated decisions that target individuals and may have legal or other types of personal consequences. However, the analysis presented here does not have any component that involves individual decision-making. It is aimed at studying the process at the population level only, just as much as and in the same way integration indices are extracted from census data. Thus, researchers

involved in any such analysis are not permitted to disclose any personal information, and all results are presented in a fully aggregated manner, in both space and time, similar to the publication of census data. Special caution should be exercised when studying smaller geographical areas: zooming in can create security concerns, so privacy evaluations should be undertaken.

Another important challenge with using social network data such as Twitter is selection bias. While Twitter is used globally, the distribution is not uniform. Some countries have broader coverage in the population compared to others, while in some countries, the platform is not used at all. Patterns of usage may also differ depending on individual socioeconomic characteristics. This means that results need to be interpreted carefully and not generalized before a proper analysis of the representativeness of the sample.

All in all, this case study shows that Twitter data can be successfully employed to study acculturation. Such analyses should be viewed as complementing and not alternate solutions to studies based on traditional data.

2.2 Measuring refugee integration through mobile phone data analysis

Mobile phones can provide detailed perspectives into the behaviour of people in a country through the mobility traces generated through their use (Ahas and Mark, 2005). Mobile phone data can also allow the gathering of insights about aspects of integration. Several ways exist to obtain useful information via mobile phones. An application installed on the phone by the user can access the GPS location (provided that the user has granted this access through the user agreement), which in turn can be used to identify home and work locations; track encounters, visited places and patterns of daily living; and know the socioeconomic status and many other indicators about the user. Given the highly sensitive character of these data, aggregation and desensitization – i.e. the removal of personal information – are essential to safeguard privacy and handle these data responsibly.

A possible source of information for measuring migrant and refugee integration are mobile phone applications specifically used by these population groups. However, the number of users and any biases in their composition must be taken into account when analysing such data. Mobile phones are like sensor packages carried by their users, and they periodically report the users' approximate locations to the telecommunications company servicing them. This is necessary to provide the communication service in the first place. One consequence is that the mobile call detail records (CDR, generated via calls and text messages) and extended detail records (XDR, also containing indications of data exchange) kept by telecommunication operators also contain traces of each user's movements. It is possible to use mobile CDR for measuring refugee integration, provided that the CDR data are properly anonymized and aggregated (Salah and Pentland et al., 2019).

The most extensive initiative linking mobile CDR to social integration has been the Data for Refugees (D4R) Challenge, which focused on Syrian refugees in Turkey and opened an anonymized data set collected from one million users over an entire year to the research community (Salah et al., 2018; Salah and Pentland et

al., 2019). The challenge prioritized five areas, of which social integration was the most popular topic among the participants.

What made D4R special among other data collaboratives using mobile CDR was that records allowed to differentiate CDR data associated (with a high probability) with refugees from other users' data. It is important to note here that Turkey is a party to the 1951 Refugee Convention, but the country recognizes "refugee" status only for people from Europe; the Syrian refugees were officially considered "temporarily protected foreign individuals". An indicator of possible refugee status was derived for customers in the database that (a) have ID numbers given to refugees and foreigners in Turkey (42.87%), (b) are registered with Syrian passports (1.06%), and (c) use special tariffs reserved for refugees (56.07%). Roughly one fifth of the database referred to such individuals.

Several indicators based on CDR analysis can be interesting from an integration perspective. Collective mobility and behavioural similarity with locals can indicate integration for various reasons (Alfeo et al., 2019). These can show the extent of interactions between locals and migrants, illustrate economic capacity for migrants, as well as capture anomalies indicating social tension.

Various aspects of migrant integration can be analysed through specific indicators (Bakker et al., 2019). The *social integration* aspects measure social ties between the migrants (or refugees) and the host country. For each migrant user, a proxy for social integration can be the number of calls that are made to native-born users relative to the total number of calls made to all users. Since an indicator is needed in the data set to compute this proxy, it is relatively rare to see this information available for analysis. The second aspect is *spatial integration*, which derives from the usage of urban spaces by migrant and non-migrant populations. Mobile CDR allows measuring proportions of these populations per area, as well as charting out of locations where native-born and migrant populations come together. For example, an index of dissimilarity can be computed for measuring spatial segregation, and an index of exposure can be calculated for measuring isolation (Iceland et al., 2002; Boy et al., 2019). Finally, *economic integration* relates to economic activities. To measure such activity, CDR data can help identify home/work locations (by the amount of activity within and outside working hours) and commuting patterns. The dissimilarity index, for example, can be computed separately for working hours and non-working hours to distinguish between residential segregation and employment segregation of the refugee and non-refugee populations in each province (Bertoli et al., 2019).

Several indicators can be directly computed via mobile data analysis. By modelling home attachment via mobile CDR, one study found that in cities where the share of refugees in the total city population is high, there is a greater tendency to interact with citizens (Boy et al., 2019). Another way to study this aspect is to find locations where refugees and locals come together (Sterly et al., 2019). Compound indicators also exist, for instance based on CDR data merged with real estate market data and media data (Bertoli et al., 2019). Extensive experiments with mobility illustrate what data sources can be used as proxies. Home-based and non-home-based activities of refugees can be analysed with "point-of-interest data" (e.g. restaurants, schools, government offices) to determine integration levels per city (Hu et al., 2019). It is important to note that home and work locations – and similarly sensitive information – are not known for

certain but calculated based on rough assumptions. Furthermore, individual-level analysis is never performed; only aggregated analysis is used, where data preprocessing ensures that no individuals can be identified or tracked using the mobile data. Ethical and privacy issues are discussed extensively in the article by Vinck et al. (2019).

Working with mobile CDR involves significant challenges. Besides difficulties in accessing such data, CDR contains data gaps. Children are mostly missing (as they are not legal owners of phone lines), and there are gender-related issues. For example, in D4R, a large proportion of the phone lines used by women are officially registered to a male (typically the husband or a son). CDR indicates presence only during communication, thus individuals who rarely communicate are less represented. This is particularly important for migrants, who may have a smaller social network to contact compared to locals.

3 Discussion and policy implications

An ideal policy approach to migrant integration should reiterate informed decision-making and public support for comprehensive and reliable statistics. Driven by known gaps in quality and compatibility of migration data (Boy et al., 2019), alternative data sources and new methodologies for developing indicators for migratory movements and migrant integration offer great potential and are increasingly being explored.

Every time a mobile device or Internet service is used by individuals, data are being generated, stored and shared by private companies. This chapter illustrated how such data can be useful for migration policymakers and discussed both the potential and the limitations of using, in particular, mobile phone data and social media data for understanding different aspects of migrant and refugee integration. We demonstrated, using Twitter as a case study, that social media platforms are a promising source for measuring different dimensions of acculturation – namely, integration, separation, marginalization and assimilation. The main benefit of this approach is that it allows expanding the geographical coverage of the analysis, providing a wide application. Our second case study focused on the uses of mobile phone data for measuring refugee integration from both social and spatial perspectives. Extensive research in ethically responsible and privacy-compliant uses of mobile CDR created new possibilities for initiation of private–public collaborations in this area (Verhulst and Young, 2019). Our case studies above illustrate that social media (Twitter) data and mobile phone data (CDR) might be used to develop indicators for home attachment and social integration. Further, each case demonstrates other potential estimates and indicators such as spatial integration and resettlement (through CDR), and different dimensions of acculturation (through Twitter data) can be developed using big data analytics. Both these data sources, once the processing pipeline is established, can provide much timelier information than traditional data-collection approaches and expand the horizon of policymakers. While CDR data offer great coverage of mobility and migrant–non-migrant encounters across a country, Twitter provides opinion and sentiment analysis opportunities, providing insights about complementary aspects.

Following validity and ethical assessments, hands-on utilization of new data sources will serve the purpose of complementing existing (traditional) data sources with timelier and more accurate estimators. While such applications can complement the shortcomings of existing (traditional) data used to inform integration policies, such as under-represented groups in migration data, interpretations are challenging since the concepts are more restricted when compared to steady and conventional methods. It is crucial to regulate how mobile phone and social media data will be processed and analysed. The existing schemes of aggregation and anonymization can ensure privacy protection and compliance with legal and ethical norms. Further collaboration between researchers, industry and policymakers on enhancing the use of artificial intelligence and machine learning for public policy decisions can prevent potential risks to fundamental rights, as described in a recent focus paper by the European Union Agency for Fundamental Rights (2018).

One of the challenges in such initiatives is that the technical concerns are often far from the policy issues, and a dialogue between data scientists and public authorities for specific cases is generally lacking (Salah and Tarik Altuncu et al., 2019). However, new initiatives on ethically responsible data science, artificial intelligence for social good initiatives, technically capable subunits of intergovernmental organizations and non-governmental organizations, and more policy-aware governance of scientific projects are helping to bridge this gap.

Acknowledgements

Authors are supported by the European Commission through the following Horizon 2020 European projects: HumMingBird – Enhanced migration measures from a multidimensional perspective (GA: 870661), SoBigData Research Infrastructure – Social Mining and Big Data Ecosystem (GA: 654024), and SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics (GA: 871042).

References

- Ahas, R. and Ü. Mark
2005 Location based services – New challenges for planning and public administration? *Futures*, 37(6):547–561.
- Alfeo, A.L., M. GCA Cimino, B. Lepri, A.S. Pentland and G. Vaglini
2019 Assessing refugees' integration via spatio-temporal similarities of mobility and calling behaviors. *IEEE Transactions on Computational Social Systems*, 6(4):726–738.
- Bakker, M.A., D.A. Piracha, P.J. Lu, K. Bejgo, M. Bahrami, Y. Leng, J. Balsa-Barreiro, J. Ricard, A.J. Morales, V.K. Singh, B. Bozkaya, S. Balcisoy and A. Pentland
2019 Measuring fine-grained multidimensional integration using mobile phone metadata: The case of Syrian refugees in Turkey. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 123–140.
- Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence and J. Weinstein
2018 Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329.
- Bertoli, S., P. Cintia, F. Giannotti, E. Madinier, C. Ozden, M. Packard, D. Pedreschi, H. Rapoport, A. Sîrbu and B. Speciale
2019 Integration of Syrian refugees: Insights from D4R, media events and housing market data. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 179–199.
- Bircan, T., D. Purkayastha, A.W. Ahmad-Yar, K. Lotter, C. Dello Iakono, D. Göler, M. Stanek, S. Yilmaz, G. Solano and Ö. Ünver
2020 Gaps in migration research: Review of migration theories and the quality and compatibility of migration data on the national and international level. HumMingBird, Leuven.
- Boy, J., D. Pastor-Escuredo, D. Macguire, R. Moreno Jimenez and M. Luengo-Oroz
2019 Towards an understanding of refugee segregation, isolation, homophily and ultimately integration in Turkey using call detail records. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 141–164.

- Calderón, C.A., F. Ortega Mohedano, M. Álvarez and M. Vicente Mariño
 2019 Distributed supervised sentiment analysis of tweets: Integrating machine learning and streaming analytics for big data challenges in communication and audience research. *Empiria: Revista de Metodología de Ciencias Sociales*, 42:113–136.
- European Union Agency for Fundamental Rights (FRA)
 2018 *#BigData: Discrimination in Data-supported Decision Making*. Vienna.
- Franklinos, L., R. Parrish, R. Burns, A. Caflisch, B. Mallick, T. Rahman, V. Routsis, A.S. López, A. Tatem and R. Trigwell
 2020 Key opportunities and challenges for the use of big data in migration research and policy. *UCL Open: Environment Preprint*.
- Hu, W., R. He, J. Cao, L. Zhang, H. Uzunalioglu, A. Akyamac and C. Phadke
 2019 Quantified understanding of Syrian refugee integration in Turkey. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 201–221.
- Iceland, J., D.H. Weinberg and E. Steinmetz
 2002 *Racial and Ethnic Residential Segregation in the United States: 1980–2000*. Volume 8. United States Government Printing Office, Washington, D.C.
- Kim, J., A. Sirbu, F. Giannotti and L. Gabrielli
 2020 Digital footprints of international migration on Twitter. In: *International Symposium on Intelligent Data Analysis*. Springer, pp. 274–286.
- Kim, J., A. Sirbu, G. Rossetti, F. Giannotti and H. Rapoport
 2021 Home and destination attachment: Study of cultural integration on Twitter. arXiv preprint arXiv:2102.11398.
- Magdy, A., T.M. Ghanem, M. Musleh and M.F. Mokbel
 2014 Exploiting geo-tagged tweets to understand localized language diversity. In: *Proceedings of Workshop on Managing and Mining Enriched Geo-spatial Data*, pp. 1–6.
- Monroe, T.
 2018 *Big Data Solutions in Forced Migration: Innovations in Analytics to Promote Humane, Sustainable Responses to Forced Migration*. World Bank Group, Washington, D.C.
- Salah, A.A., M. Tarik Altuncu, S. Balcisoy, E. Frydenlund, M. Mamei, M. Ali Akyol, K. Yavuz Arslanlı, I. Bensason, C. Boshuijzen-van Burken, P. Bosetti, J. Boy, T. Bozcaga, S. Mümin Cilasun, O. Işık, S. Kalaycıoğlu, A. Seyyide Kaptaner, I. Kayı, Ö. Ozan Kılıç, B. Kjamili, H. Kucukali, A. Martin, M. Lippi, F. Pancotto, D. Rhoads, N. Sevenscan, E.

- Sezgin, A. Solé-Ribalta, H. Sterly, E. Surer, T. Taşkaya Temizel, S. Tümen and I. Uluturk
 2019 Policy implications of the D4R challenge. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 477–495.
- Salah, A.A., A. Pentland, B. Lepri and E. Letouzé (eds.)
 2019 *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer Nature Switzerland AG, Cham.
- Salah, A.A., A. Pentland, B. Lepri, E. Letouzé, P. Vinck, Y.A. de Montjoye, X. Dong and O. Dagdelen
 2018 Data for refugees: The D4R challenge on mobility of Syrian refugees in Turkey. arXiv preprint arXiv:1807.00523.
- Sírbu, A., G. Andrienko, N. Andrienko, C. Boldrini, M. Conti, F. Giannotti, R. Guidotti, S. Bertoli, J. Kim, C.I. Muntean, L. Pappalardo, A. Passarella, D. Pedreschi, L. Pollacci, F. Pratesi and R. Sharma
 2021 Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11:341–360.
- Sterly, H., B. Etzold, L. Wirkus, P. Sakdapolrak, J. Schewe, C.F. Schleussner and B. Hennig
 2019 Assessing refugees' onward mobility with mobile phone data – A case study of (Syrian) refugees in Turkey. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 251–263.
- Verhulst, S.G. and A. Young
 2019 The potential and practice of data collaboratives for migration. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 465–476.
- Vinck, P., P.N. Pham and A.A. Salah
 2019 “Do no harm” in the age of big data: Data, ethics, and the refugees. In: *Guide to Mobile Data Analytics in Refugee Scenarios* (A.A. Salah, A. Pentland, B. Lepri and E. Letouzé, eds.). Springer Nature Switzerland AG, Cham, pp. 87–99.
- Wimark, T., K. Haandrikman and M.M. Nielsen
 2019 Migrant labour market integration: The association between initial settlement and subsequent employment and income among migrants. *Geografiska Annaler: Series B, Human Geography*, 101(2):118–137.
- Berry, J.W.
 2019 *Acculturation: A Personal Journey across Cultures*. Cambridge University Press.