



HAL
open science

The Empirical Study of Human Mobility: Potentials and Pitfalls of Using Traditional and Digital Data

Ettore Recchi, Katharina Tittel

► **To cite this version:**

Ettore Recchi, Katharina Tittel. The Empirical Study of Human Mobility: Potentials and Pitfalls of Using Traditional and Digital Data. Handbook of Computational Social Science for Policy (ed. Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, Michele Vespe), Springer International Publishing, pp.437-464, 2023, 9783031166242. 10.1007/978-3-031-16624-2_23. hal-03956897

HAL Id: hal-03956897

<https://hal-sciencespo.archives-ouvertes.fr/hal-03956897>

Submitted on 25 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

Chapter 23

The Empirical Study of Human Mobility: Potentials and Pitfalls of Using Traditional and Digital Data



Ettore Recchi and Katharina Tittel

Abstract The digitization of human mobility research data and methods can temper some shortcomings of traditional approaches, particularly when more detailed or timelier data is needed to better address policy issues. We critically review the capacity of non-traditional data sources in terms of accessibility, availability, populations covered, geographical scope, representativeness bias and sensitivity, with special regard to policy purposes. We highlight how digital traces about human mobility can assist policy-making in relation to issues such as health or the environment differently to migration policy, where digital data can lead to stereotyped categorizations, unless analysis is carefully tailored to account for people's real needs. In a world where people move for myriad reasons and these reasons may vary quickly without being incorporated in digital traces, we encourage researchers to constantly assess if what is being measured reflects the social phenomenon that the measurement is intended to capture and avoids rendering people visible in ways that are damaging to their rights and freedoms.

23.1 Introduction

Besides the shock of the human lives lost to the disease, when the Covid-19 pandemic broke out, the global public opinion was in awe by the sight of the main non-pharmaceutical intervention that governments put in place almost everywhere – lockdowns. The images of an immobile world created an unexpected dystopic landscape. The prohibition to step out of home emptied cities, highways, stations and airports. Like in a postatomic fantasy, media spread out pictures and videos of the usually most crowded venues of the planet – from Shibuya Station in Tokyo to

E. Recchi (✉) · K. Tittel

Sciences Po, Centre de Recherche sur les Inégalités Sociales (CRIS), CNRS, Paris, France

Migration Policy Centre (MPC), EUI, Florence, Italy Institut Convergences Migrations, Paris, France

e-mail: ettore.recchi@sciencespo.fr; katharina.tittel@sciencespo.fr

© The Author(s) 2023

E. Bertoni et al. (eds.), *Handbook of Computational Social Science for Policy*,
https://doi.org/10.1007/978-3-031-16624-2_23

437

Heathrow Airport in London – without a living soul in the middle of the day. After decades of reckless increase in the number of airline travellers, for instance, in April 2020 these were no more than 3 per cent their number a year earlier (Recchi et al., 2022). No other nonviolent event could have looked more antinomic – and thus revealing – of the nature of social life in late modernity.

In the pre-Covid-19 era, the number of international trips had been on a steady and uninterrupted rise since at least 1960 – from 69 million to almost 3 billion border crossings per year (Recchi, 2015, 2016; Recchi et al., 2019). Never in history have human beings had such an ease to move out of their usual residences – whether on daily commutes, weekend trips, holiday travel or (although not across the board) long-term migration spells.¹ This is clearly more the case for the rich, whose lifestyle is often patterned after frequent journeys. Nonetheless, the drop in the cost of travel has also facilitated the (shorter-haul and less exotic) mobility of the less privileged, at least in high-income countries (e.g. Demoli and Subtil, 2019). Spatial mobility has thus progressively become a hallmark of our age, as social theorists Zygmunt Bauman and John Urry happened to remark already by the turn of the millennium (Bauman, 1998; Urry, 2000). In the second half of the twentieth century, the absence of large-scale wars, economic growth, progress in transportation and ICT developments – that is, the major keys to globalization – paved the way to a more mobile world. While the aftermath of the Covid-19 pandemic and a possibly higher sensitivity to the climate impact of fuel-propelled mobility may stymie this pre-existing trend, human mobility will hardly cease to be part and parcel of what it means to live in the twenty-first century – not the least because global inequality and climate change also spur migration (Barnett & McMichael, 2018; Milanović, 2019; Rigaud et al., 2018).

For a comprehensive take of the topic, we must acknowledge its different manifestations – first of all, in spatial terms. Human mobility is a multiscale phenomenon, spanning from the micro (local) to the meso (national) and macro (international) level. Repeated national surveys show that all of these levels saw an increase in the last decades of the twentieth century (for instance, in Germany: Zumkeller, 2009). People spend more time in mobility and cover a larger distance per time/unit than they were used to do. The second dimension that thus needs to be acknowledged is the temporal manifestation of mobility. Movements can be temporary, seasonal or longer term/permanent. The third dimension that fundamentally shapes individuals' mobility experiences as well as legal and policy responses to it is the reason for mobility, including voluntary reasons (such as tourism, work, education or family reasons) or forced displacement (as a result of conflict or natural disasters), with or

¹ While short-term mobility has skyrocketed in comparative terms, international migration only modestly increased in recent decades: in 2019, 3.5% of the global population qualified as migrants compared to 2.9% in 1990 (UNDESA, 2019). This proportion is obtained applying the UN definition, according to which an international migrant is somebody who has moved into a different country for 12 months or more (see the Glossary on Migration in IOM 2019). Issues of definition are discussed in the draft *Handbook on Measuring International Migration through Population Censuses* (UNDESA, 2017).

without documentation. In practice demarcations are not always clear, since reasons are often mixed and people are often motivated to move by a multiplicity of factors (Mixed Migration Centre, 2020), and intersections between migration and other forms of circular mobility are growing (Skeldon, 2018). Think, for instance, about ‘gap years’ and ‘sabbaticals’, not only in academia.

While mobility is a fundamental element of human freedom with real and perceived value for all groups affected by it, the historical intensification of mobility – and its likely spatial-temporal clustering in certain sites (for instance, global cities) or certain periods (for instance, during end-of-year festivities) – is a key concern from a policy perspective. We mention here just two issues that have come to the fore in recent years, partly in conjunction with the Covid-19 crisis – beyond the epidemiological risk that is openly associated with the movement of virus-carrying population.

First is the ‘evasiveness of remote workers’ through mobility. Since the outbreak of the pandemic, firms increasingly operate partly or completely remote, asking or allowing workers to ‘work from home’. Such a shift, which was already present in some industries like IT, may spur white collars’ relocation even at long distance – including abroad – as ‘digital nomads’. Some countries introduced so-called digital nomad visas (Bloom, 2020; Hughes, 2018) that can be used to track such mobility. Mobility in free movement zones, or if workers relocate short term on tourist visas, often goes unrecorded yet may be of increasing interest for policies related to such mobilities, including tax issues.

Second is the ‘environmental and epidemiological risks’ linked to human mobility. The bulk of human mobility is fed by fossil fuel burned for road, rail, air and maritime transportation. Almost all (95 per cent) of the world’s transportation energy comes from petroleum-based fuels, largely gasoline and diesel. Globally, transportation accounts for no more than 14 per cent of the global greenhouse emissions – less than electricity and heating (25 per cent), agriculture (24 per cent) and industry (21 per cent) (Pachauri et al., 2014). These proportions vary significantly by level of economic development though, and transportation takes the lion’s share of greenhouse gas emission in richer countries (e.g. the US: EPA, 2019). Therefore, the impact of mobility on the environment is bound to become more severe as economic development advances, unless major changes in fuel emission take place. In parallel, travel spreads diseases, and increased travel may have made the world more vulnerable to epidemics, although the intensity of long-distance mobility does not necessarily entail a stronger incidence of epidemics (Clemens & Ginn, 2020; Recchi et al., 2022).

The importance of effective policy responses to mobility has been amplified during the Covid-19 pandemic. A major puzzle is whether human mobility will change in size, scope and form in the coming years. At the macroscale, the appetite of human beings for travel does not seem shaken, but travel limitations and travellers’ biometric and health controls are likely to be enhanced as ‘new normal’ ways of restricting (even surreptitiously) access to undesirable travellers (Favell & Recchi, 2020). At the meso and micro level, it is unlikely that the economic and cultural attractiveness of cities as poles of mobility will be disrupted, although the

take-off of telework – as complement or substitute to office spaces – may incentivize some sort of ‘flight to the suburbs’ (Florida et al., 2021). Clearly, the pandemic moment and its aftermath expose the importance of monitoring human mobility with adequate measurement tools to improve response capacities. At the same time, the pandemic experience has also brought more attention to the ethical components of tracking mobility among the general public.

23.2 Monitoring Human Mobility: Traditional and New Data

Tracking human movements in space has always been a challenge for population statistics, but new data open up new opportunities and challenges. Previous studies (European Commission, 2016; Bosco et al., 2022) offer a systematic review of the literature about measuring migration with traditional and new data sources, which we complement with up-to-date information and critical consideration from a policy-related angle. As suggested also by Taylor in this volume (Taylor, 2023), we pay particular attention to what digital data reflects and how it can be used for policy-relevant analysis, foregrounding the policy issue to be solved and what evidence is needed in support, discarding the ‘panopticon illusion’ (and danger) of making everything visible through mobility data.

23.2.1 *Traditional Data: Pros and Cons*

Traditionally, mobility has been measured with censuses, population registers, administrative sources and household surveys. Data from these sources are cleaned, edited, imputed, aggregated and used to produce official statistics, including the datasets documenting international migration flows and migrant stocks released by agencies of the United Nations.

The major advantages of traditional data sources are that they are transparent, frequently curated and stored in public databases (with varied degrees of accessibility), allowing comparability over time and across countries. However, there are some important limitations. First, they are not reliably available in many parts of the world. Estimates on in- and out-migration flows by country of origin and destination are only reported by 45 countries to the United Nations (UN DESA, 2015). Second, these aggregate statistics have a poor time and spatial resolution. Moreover, with these standard approaches, the category of internally displaced persons is often overlooked despite its policy relevance. In general, inconsistent definitions of a *migrant* make it difficult to compare data across different countries (Sîrbu et al., 2021). Third, and as an extension of the previous point, traditional sources typically do not capture circular, short-term, seasonal or temporary mobility (Hannam et al., 2006). Fourth, surveys that include large enough samples of people with different migratory backgrounds and socioeconomic profiles, particularly the most

vulnerable, in different contexts and over time are not at all or not systematically available despite the importance to understand inequalities in terms of education, housing, employment, discrimination, well-being, access to services and protection, etc. Finally, censuses and surveys have data publishing lags of several years.² This is particularly problematic in a context in which migratory flows become increasingly complex and dynamic, and in emergency situations, including environmental or health crisis situations.

Both academic and nonacademic actors have tried to improve comparability and availability of traditional data sources and reconcile measurement problems, such as undercount, varying duration of stay criteria and coverage (de Beer et al., 2010; European Commission, 2016; Raymer et al., 2013). To respond to these shortcomings, and with the purpose of informing the humanitarian community and government partners, different international organizations have established data collection and dissemination mechanisms on specific aspects of human mobility, such as UNHCR's refugee statistics, ILO's labour migration statistics, the World Bank remittances database or IOM's data on various migrations matters including internal displacement and their 'missing migrants' project. These organizations are increasingly aiming to incorporate more digital, non-traditional data sources as part of their migration data strategies.

23.2.2 Non-traditional Data Usages: An Overview

The mass use of digital devices across the globe has generated large repositories of spatiotemporal 'trace data' (Chi et al., 2020), some of which provide new opportunities for ad hoc measurements and modelling of human mobility. While new technologies are capturing mobility rather than migration data (McAuliffe & Sawyer, 2021), some can also be used to better understand certain aspects of migration. As outlined in Table 23.1, different non-traditional data sources differ significantly in terms of the information available, the populations covered, geographical availability, the data level (individual or grouped), representativeness bias issues, sensitivity and in consequence in terms of who they reflect, the mobility events they capture (micro, meso and macro level), ethical issues and their usefulness to provide information relevant for policy purposes. In the policy sphere, categories are used to define 'groups of people who are assumed to share particular qualities that make it reasonable to subject them to the same outcomes of policy' (Bakewell, 2008: 436). While in relation to issues such as health or the environment, information about mobility events (how many individuals move, where, when and how) provide key information and the characteristics of who moves may be secondary, in the context of migration, analytical or administrative categories, such

² For example, in the case of the International Migration Database of the OECD, the lag is between 2 and 3 years.

Table 23.1 Characteristics of some traditional and non-traditional data sources for the empirical study of human mobility

	Data source	Populations covered				Accessibility and costs				
		Information ^a	Geographical availability	Data level ^b	Time invariance ^c	Representativeness bias issues ^d	Data costs	Accessibility	Frequency of collection	Release time ^e
Traditional data sources	Censuses	Country of birth, citizenship, year of arrival in the country for foreign-born persons, place of residence (or change of)	Absent or less reliable in low- and middle-income countries	Country level	Yes	Biased toward documented citizens, often failing to cover undocumented migrants	Expensive, time consuming	Usually public	Every 5 or 10 years	Delayed
	Population registers	Inventory of the population living in a particular area	Absent or less reliable in low- and middle-income countries	Local or country level	Yes	Emigrants tend to be under-counted lacking incentives to deregister	Cheap, research data is by-product	Usually made accessible for research purposes	In real time (updated following vital events)	Delayed

	Administrative data	Data retrieved from issuance of residence permits, border statistics and health insurance registrations	Absent or less reliable in low- and middle-income countries	Local or country level	Yes	Not capturing free or irregular mobility	Cheap, research data is by-product	Usually made accessible for research purposes	In real time (updated following vital events)	Delayed
	Surveys	Targeted information, also retrospective panel-like	Exceptional in low- and middle-income countries	Country level, some coordinated international-ally	In theory, but few longitudinal surveys	Inferential problems with small minorities in sample surveys	Expensive, time consuming	Usually public	Sporadically	Delayed
Non-traditional data sources	Geotagged CDR	Individual ID, timestamps and geo-coordinates from each activity (e.g. calling)	Global, but use and ownership vary across countries and over time	Country level, some coordinated international-ally	No, but population relatively constant	Likely and hard to assess, depending on phone use patterns	Cheap, research data is by-product	Proprietary, requires purchase agreements	In real time	Depends on purchase agreement

(continued)

Table 23.1 (continued)

		Populations covered				Accessibility and costs			
Data source	Information ^a	Geographical availability	Data level ^b	Time invariance ^c	Representativeness bias issues ^d	Data costs	Accessibility	Frequency of collection	Release time ^e
Twitter	Individual ID, timestamps and geo-coordinates from each activity (e.g. posting)	Global, but adoption rate very different by country	Global	No, population changes rather quickly	Likely and hard to assess, biased toward educated, young, politically active, exacerbated by bots	Cheap, research data is by-product	Through API	In real time	Instantly
Facebook ads	Stock estimates and profiles of specific populations also for non-probability sampling	Global, but adoption rate very different by country	Global	No, population changes rather quickly	Biased toward younger and middle-aged individuals	Cheap, research data is by-product	Through API	In real time	Instantly
Satellite data	Images	Global, but can detect mobility only in certain instances	Depends on local context					In real time	Instantly

Google/Apple mobility datasets	Pre-calculated indexes that show intensity by different transport types	Global, but adoption rate very different by country	Global	No, population tends to change rather quickly	Likely and hard to assess as companies do not share the characteristics of users	Cheap, research data is by-product	Limited by company	In real time	Depends on company choices
Flickr, Instagram	Geolocation of photos and text	Mostly in high- and middle-income countries, but adoption rate differs by country	Global	No, population change rather quickly	Biased toward younger, female users	Cheap, research data is by-product	Through API	In real time	Instantly
LinkedIn	Workplace, occupational sector, languages spoken, education	Mostly in high- and middle-income countries, but adoption rate differs by country	Global	No, population tends to change rather quickly	Biased toward highly educated, younger workers	Cheap, research data is by-product	Through API	In real time	Instantly
Flight data	Numbers of passengers and routes	Global	Global	Potentially longitudinal data but unreleased	No bias, captures all travellers	Cheap, research data is by-product	Subject to (purchase) agreements with firms	In real time	Depends on purchase agreement
Ticket reservation data in bus, railway or sea lines	Numbers of passengers and routes	Mostly available in high- and middle-income countries	Local, national and international, but not integrated	Potentially longitudinal data but unreleased	No bias, captures all travellers	Cheap, research data is by-product	Subject to (purchase) agreements with firms	In real time	Depends on purchase agreement

(continued)

Table 23.1 (continued)

Traditional data sources	Mobility events captured			Types of mobility captured		Risk for privacy and confidentiality				
	Data source	Micro	Meso	Macro	Information on migrant status ^f	Information on short-term/circular migration ^g	Nature of data ^h	Informed consent	Confidentiality risk ⁱ	Vulnerable population
Censuses	No	Imprecise (can be inferred from residence change)	Imprecise (can be inferred from residence change)	Imprecise (can be inferred from residence change)	Yes (imperfect, based on, e.g. place of birth)	No	Individual	Yes	Low (informed consent)	No
Population registers	Yes (if involving change of residence)	Partly (in countries where individuals have to register their residence)	Partly (in countries where individuals have to register their residence)	Incoming migration captured; outgoing often is not	Yes (imperfect, based on, e.g. place of birth)	No	Individual	Yes	Low (informed consent)	No
Administrative data	No	Partly (in countries where individuals have to register their residence)	Partly (in countries where individuals have to register their residence)	Partly (e.g. for visas)	Yes	No	Individual	Yes	Low (informed consent)	No
Surveys	No	Partly	Partly	Partly	Yes (depends on the survey)	No	Individual	Yes	Low (informed consent)	No

Non-traditional data sources	Geotagged CDR	Yes	Yes	Depends on context	No (exception: Data4 Development challenge)	Depends on context	Individual	No	Very high (re-identification and linkage to sensitive places)	Yes
	Twitter	Yes (limited)	Yes (limited)	Yes (limited)	Possibly (imperfect, based on language or network properties)	Yes (in theory)	Individual	No	High (unobtrusively collected on the level of the individual)	Depends
	Facebook ads	No	No	No	Estimated by Facebook, methodology not transparent	No	Aggregate	No	Low (either just aggregated data or further surveys with informed consent)	Depends
	Satellite data	Yes	No	No	Yes (possibly to detect displaced populations)	Depends on context	Aggregate	No	Occasionally (i.e. population movements in contested areas)	Yes

(continued)

Table 23.1 (continued)

Data source	Mobility events captured			Types of mobility captured		Risk for privacy and confidentiality			
	Micro	Meso	Macro	Information on migrant status ^f	Information on short-term/circular migration ^g	Nature of data ^h	Informed consent	Confidentiality risk ⁱ	Vulnerable population
Google/Apple mobility datasets	Yes (limited)	No	No	No	No	Aggregate	No	Low (unobtrusively collected but only shared in aggregated form)	No
Flickr, Instagram	Yes (limited)	Limited	Limited	Possibly (imperfect, based on language or network properties)	Yes (in theory)	Individual	No	Medium (unobtrusively collected at individual level but information relatively less sensitive)	Depends

	LinkedIn	No	Yes (limited)	Yes (limited)	Possibly (imperfect, based on language or education)	No	Individual	No	Medium (unobtrusively collected at individual level but information relatively less sensitive)	No
	Flight data	No	In case of internal flights	Yes	No	No	Aggregate	No	Low (unobtrusively collected but low risk of identification)	No
	Ticket reservation data in bus, railway or sea lines	Depends on type of transport	Depends on type of transport	Depends on type of transport	No	No	Aggregate	No	Low (unobtrusively collected but low risk of identification)	No

^aWhich information is usually used to detect mobility events or migrant identity?

^bAt which level is data usually collected?

^cIs the population relatively constant over time?

^dIs there a systematic bias in the sample?

^eWhen is the data usually made available?

^fCan data be used to infer migrant status?

^gCan short-term or circular migration be measured?

^hFor research purposes, is data usually shared at the level of individuals or as aggregate?

ⁱIs there a risk of re-identification of the individual?

as ‘migrant’, ‘foreign worker’, ‘internally displaced person’ or ‘refugee’, fundamentally shape the interactions between individuals and bureaucratic organizations. As Taylor stressed in this volume (Taylor, 2023), that connection is often obscured when computational methods and new data sources are used.

23.2.2.1 A Review of the Usefulness of Non-traditional Data to Study Different Types of Mobility

Local and National Mobility

At micro and meso level, geotagged digital trace data from call detail records (CDR) (e.g. Song et al., 2010), GPS technology (e.g. Bachir et al., 2019; Cui et al., 2018; Huang et al., 2018) or social media data (e.g. Bao et al., 2016) can be used to study individual (Giannotti et al., 2011; González et al., 2008; Pappalardo et al., 2015; Wang et al., 2011) as well as group mobility (Hir et al., 2019; Lulli et al., 2017; Tosi, 2017). Because of their wide coverage³ and ad hoc availability, these data allow studying population movements in emergency situations, such as during natural disasters (Bengtsson et al., 2011) or events like the Covid-19 pandemic (e.g. Xiong et al., 2020). In other contexts, satellite data have been used to estimate the effect of extreme climate events, such as flooding, on migration (Chen et al., 2017). Compared to self-reporting on causes of migration in surveys, they offer the advantage of not being affected by subjective factors such as recall bias.

While individual characteristics, such as gender or age or motivations for mobility, that are key variables to consider for policy responses, are usually unavailable, researchers started to collect or link survey data with geotagged digital trace data to alleviate this limitation and get more information on demographic characteristics of the populations covered (Blumenstock & Fratamico, 2013). Other sources used for mobility research include Twitter (e.g. Fiorio et al., 2017; Zagheni et al., 2014), Skype (e.g. Kikas et al., 2015), LinkedIn (e.g. Li et al., 2019) or Flickr (Bojic et al., 2016) and could include any other platform that provides geotagged data of their users. Their usefulness for policy purposes fundamentally depends on how well represented the population of interest is on the specific platform.

Large platform companies like Apple or Google also possess vast repositories of human movement data that could be used to understand local mobility patterns. While these companies do not normally publish their data for research purposes, they offered ad hoc data products and visualizations of aggregated mobility of customers, including the use of travel modes (public transport, driving, walking), during the Covid-19 pandemic (Apple, 2021; Google, 2021). Notably, however, omitted information on methods and on the underlying population that is captured

³ As of 2018, mobile phone penetration is around 100 per cent in high- and middle-income and 55 per cent in low-income countries. This has raised from 12 per cent of the world population in 2000 (Worldbank, 2019).

leads to a lack of clarity of these data and their biases, limiting their usefulness for policy purposes.

International Mobility

Social media advertising platforms can help estimate stocks and sociodemographic profiles of certain populations and facilitate non-probability sampling, since the platforms support showing ads exclusively to certain audiences. This information has also been used to target specific populations in order to invite them through paid Facebook advertisement to participate in a survey, such as Polish migrants in European countries (Pötzschke & Braun, 2017). Compared to traditional surveys, this approach offers the advantage of targeting demographic characteristics to reach a larger sample size at a global scale quickly and at lower cost (Rampazzo et al., 2021).

Böhme et al. (2020) used georeferenced online search data from Google Trends (looking for the combination of migration- and target country-related keywords as a proxy for migration intentions) in origin countries to improve the predictive power of international migration models. While there are promising examples in different areas employing Google Trends data, such as to forecast private consumption (Vosen & Schmidt, 2011), the precision and goodness of fit of such models can also rapidly change (Lazer et al., 2014).

A Special Case: Airline Mobility

A major source of big data on travel are airline reservation systems (ARS). A handful of private companies dominate this market. They handle such information omitting not only personal information but also categorical groupings about sociodemographic characteristics of passengers. One of these companies, Sabre, sells an air travel dataset that reports monthly data on the numbers of air travellers between all world airports and regular airline routes. Capitalizing on this source, in combination with the more traditional statistical reports of the United Nations World Tourism Organization, researchers have created a Global Transnational Mobility Dataset which details cross-border trips between all sovereign states worldwide from 2011 to 2016 (Recchi et al., 2019). Other studies have used Sabre data to infer types of transnational mobility (Gabrielli et al., 2019) the economic impact of reduced mobility due to Covid-19 (Iacus et al., 2020), and the global spread of the pandemic in 2020 (Recchi et al., 2022).

Potentially, similar data could be collected for other ticket reservation systems in bus, railway or sea lines, but such forms of transportations tend to be highly national or regional, rather than global, and thus there is possibly an issue of integration of different sources. At any rate, this is an evolving area of data collection that has proven fruitful for macro analyses of international flows. Its major limit is that a travel is an event, not a person, thus leaving uncharted the characteristics of human

populations that experience cross-border travel, which survey research describes as mostly – albeit not exclusively – drawn from among the middle-upper classes (Demoli & Subtil, 2019).

Along these lines, Chareyron et al. (2021) used data scraped from the digital platform Tripadvisor to examine privileged mobility patterns. Other platforms for evaluating tourism consumption (accommodation, places, activities) that might be leveraged for research on this issue in certain contexts include Booking, Airbnb, [Hotels.com](https://www.hotels.com) or Weibo.

Difficulties to Infer Policy-Relevant Categories from Digital Trace Data

While digital devices trace the geolocalizations of their users, there are no standards or commonly respected methodological frameworks for how to produce estimates of policy-relevant information from granular geo-located data points (Bell et al., 2015), and the analysis of such data by data scientists without context-specific knowledge and understanding of the social phenomena underlying human mobility creates new risks (McAuliffe & Sawyer, 2021). Unlike survey data about respondents' residential history, georeferenced digital trace data only record locations at a specific moment in time. Blondel et al. (2015) and Chi et al. (2020) introduce different estimation techniques to infer patterns of human mobility from observational geo-tagged data. Without further context-specific information, it is not straightforward to determine what the location of a given individual corresponds to (Fiorio et al., 2021). For this reason, how researchers choose to define features of trips for the ambiguous distinction between migration and other kinds of movements, and how they group geo-located data points together based on their temporality, greatly affects the consistency of human mobility estimates generated from digital trace data (Ahas et al., 2018; Fiorio et al., 2021). This points to the challenge of how to discern policy-relevant categories from inferred mobility patterns. A risk that is linked to this labelling process is called *delinkage*, which refers to the replacement of an individual identity by a 'stereotyped identity with a categorical prescription of assumed needs' (Zetter, 1991: 44).

In theory, the possibilities opened by the new data sources suggest revisiting some presuppositions of such labelling and categorization processes and question the labels researchers apply to people and the functions those categories fulfil to design policies that effectively cater for real and not stereotyped human needs (see Turton, 2005; Bakewell, 2008). In practice, however, there are several challenges to correctly understand and interpret the underlying meaning of data variables in different contexts. Certain data, such as those scraped from LinkedIn, may offer relatively straightforward ways to identify a 'foreign worker' on the platform, whereas such classification is more difficult and more sensitive to contextual changes when using CDR data. Ahas et al. (2018), for example, in their roaming dataset operationalize a 'foreign worker' as someone who did 1 to 52 trips to a certain country in a certain time period. Such an identification strategy would not have worked, however, during the Covid-19 pandemic, when remote work was widespread. Geolocated messages

or posts are often the key variable in estimating the geo-coordinates of users, while other studies use the language used on social networks; friend or follower networks; profile pictures; names; or other textual information available (e.g. Huang et al., 2014; Kim et al., 2020) to infer users' sociodemographic characteristics. In the case of Facebook marketing data, researchers have to rely on the categories provided by the platform, even though, as Zagheni et al. (2017) highlight, categories are not documented according to scientific research standards. This may introduce biases that are hard to disentangle from biases related to selection and non-representativeness, or other inconsistencies. 'Naming' mobile individuals is often based on legal definitions and should be carefully considered, particularly in a context where inaccurate estimates can cause confusion and be fuel for heavily contested public and political discourses (McAuliffe & Sawyer, 2021).

Without relevant content knowledge of migration and technology use, errors or wrong assumptions can lead to misspecification and misinterpretation (McAuliffe & Sawyer, 2021), exemplified by Pew Research's 2019 estimates of irregular migrants in Europe (Connor & Passel, 2019). The authors wrongly included asylum seekers whose applications were being processed in the category of irregular migrants, leading to inaccurate and inflated estimates. Drawing on examples like this one, McAuliffe and Sawyer (2021) highlight that, in reality, the application of so-called new data science in the study of migration often fails to take into account the most basic understanding of the topic.

23.2.2.2 Limitations and Caveats in the Use of Non-traditional Data on Human Mobility

Despite the opportunities offered by non-traditional data sources, their use comes with important limitations and caveats that add to the potentials we outlined so far. In this final section, we list and discuss four of them. Importantly, the different non-traditional data sources differ significantly with regard to their properties and hence related concerns.

Proprietariness

A first, and rather mundane, problem with some of the above-mentioned data is difficult access. Some data sources require appropriate technical skills (e.g. Facebook and LinkedIn marketing API); some data can be purchased; some sources lack formalized purchasing mechanisms (e.g. mobile phone providers), and others do not share their data at all. Moreover, by employing terms of service (TOS)-compliant methods, a researcher may respect the business prerogatives of the company that created the platform studied, but this may or may not respect the dignity and privacy of the platform users (Freelon, 2018). This is particularly sensitive in a context of radical power asymmetries with the platform/service providers, as users often have far less understanding of who can access their data

and under which circumstances, as well as of the functioning of the tools they use online (Broeders & Dijstelbloem, 2015; Taylor, 2023).

Non-representativeness

Second, another key and too often overlooked issue with digital trace data – like in many other social science data – is selection bias: users of a particular social media platform or mobile phone provider are not representative of the underlying general population. In the analysis of CDR, selection bias regarding mobile phone ownership and usage must be considered when extrapolating from the number of moving SIM cards to the number of moving persons (Blumenstock, 2012; Blumenstock & Fratamico, 2013). For instance, in some sub-Saharan African countries, men are more likely to be mobile phone owners, while phone sharing is common among rural women, and there is considerable cross-country variation: while mobile phone records in Kenya are an excellent proxy for mobility, regardless of socioeconomic factors, mobile phone data in Rwanda are a good proxy only for the mobility of wealthy and educated men (Luca et al., 2021). While existing studies showed that approaches using CDR data work well in one-off emergencies, such as the earthquake in Haiti (Bengtsson et al., 2011) and other disaster events (Chen et al., 2017), for estimating general population displacement, ad hoc knowledge is needed about who is using phones or services. Otherwise, such approaches cannot identify vulnerabilities of specific populations, a key aspect of targeting social protection and relief (Lu et al., 2016). Similarly, Facebook and Twitter adoption rates differ between countries and depending on user characteristics, such as age or gender (Zagheni et al., 2017). By relying on data from highly specialized online services, users' self-selection into these services hence limits the generalization of these results (Böhme et al., 2020). For instance, LinkedIn may be useful to study the labour mobility of highly educated individuals in rich countries and allows researchers to link this to career choices and industry-specific patterns. However, it cannot yield mobility estimates for the global population. This is problematic because, as Sirbu et al. (2021) highlight, being unable to track specific groups of users can steer migration policies in directions that unwillingly perpetuate discriminations or neglect the needs of invisible groups.

Different statistical approaches help to correct for selection bias. Zagheni and Weber (2015) propose a method that relies on calibration of the digital trace data against reliable official statistics. When the data also contains demographic information about users of a given platform, that information can be leveraged to de-bias non-representative results by adjusting the responses via multilevel regression prediction models and post-stratification (Wang et al., 2015). Importantly, statistical calibration models require datasets containing enough variables for the use of post-stratification techniques, as well as knowledge about specific functional relationship between estimates of migration and how this relationship varies by geography and population characteristics as well as how it changes over time – information that is often not available, and that requires systematic and hence costly on-the-ground

research. Since the composition of the user bases of new data sources may change rapidly, predicting over time variation is usually more difficult than understanding cross-spatial variation in human mobility.

Beyond challenges that are common to any survey, such as selection bias and nonresponse, some pitfalls are specific to non-probability sampling on social media. Not only does non-probability inclusion lead to non-representative data, but the sampling error is further enhanced by the self-selection of users, which may be affected by issues of trust and incentives, and by the platform's algorithm.

To alleviate some of these shortcomings and generate more reliable and comprehensive estimates, it is key to borrow from a number of different data sources and develop methods to analyse them that are robust to the lack of a specific data source (European Commission, 2016). For example, Huang et al. (2021) used Twitter, Google, Apple and Descartes Labs data to disentangle the disparities in mobility dynamics from lower- and upper-income US counties during Covid-19. They found that mobility from each source presented unique and even contrasting characteristics. Their (optimistic) conclusion is that hierarchical Bayesian methods can be used effectively to combine different mobility data in a consistent way. However, this requires the availability of different mobility datasets as proxies for the same phenomenon, which at the global level – given that the data showed contrasting characteristics even for the USA – seems extremely hard to achieve.

No Gold Standard

Third, it must be acknowledged that a proper gold standard does not exist since precise current and past mobility patterns are unknown. Therefore, validation of nowcasting models of human mobility is not straightforward. While traditional data sources have a number of limitations and caveats, without a benchmark it is difficult to trust new data sources and innovative approaches and assess their validity (European Commission, 2016). Therefore, a combination of traditional and new data might yield more accurate estimates and predictions than solely relying on non-representative sources (Lazer et al., 2014; Zagheni et al., 2017).

Ethical Concerns

Finally, we deem appropriate to underscore some ethical caveats and raise frequently ignored data justice-related questions, which Taylor discusses in this issue (Taylor, 2023). While statistical techniques may alleviate the shortcomings of new data sources, the use of some of this information, notably individual-level data (CDR, social media data), raises severe ethical issues. Anonymization – i.e. removing personal identifiers – is a commonly used method to protect users' privacy, but it is not sufficient to shield privacy nor address issues related to informed consent, since in large mobility datasets, individuals can be reidentified with as little as four spatial-temporal data points, even if they do not contain identifiable information like names

or email addresses (de Montjoye et al., 2013). Having a precise, always-on tracking of individuals, with a spatiotemporal history of their trajectories, and drawing a picture of how people use city space or move across borders and how they break rules and create informal ways to support themselves are a sensitive matter in any context, especially when data are unobtrusively collected without informed consent. Risks are aggravated in an environment where geolocations might be mapped to addresses such as religious places, abortion clinics and other sensitive areas. In the context of migration, where many individuals are vulnerable, and political freedoms cannot be taken for granted, these concerns are particularly important. It is key to consider what it means if mobile populations become more legible and, thereby, more amenable to control from above (Scott, 2008). Individual invisibility may sometimes be life-saving or, at least, grant a basic right to personal freedom. As Polzer and Hammond (2008) insist, ‘researchers who lift this veil [of invisibility] in the name of illuminating ‘creative livelihood strategies’ or ‘flexible identities’ may inadvertently be alerting powerful states, the UN or NGOs to the ways in which their rules are circumvented, and thereby reduce the space for life-saving creativity and flexibility in remaining invisible’. While visibility to institutions that are seen as potential allies might increase access to resources, defence of rights and legitimacy and can hence be seen as an ethical imperative, invisibility may serve as a protective shield in the absence of true legal, political and social protection, and in contexts of xenophobic and majoritarian violence. As governments and public agencies are increasingly using digital technologies for a more efficient, neutral and disembodied migration management and border control (Latonero & Kift, 2018; Trimikliniotis et al., 2015), Leurs and Smets (2018) remind us that it is important to ponder how approaches, methodologies, tools and findings may be coopted or used in unintended and undesirable ways. Actors interested in better understanding human mobility may include organizations like the United Nations and aid agencies, but also private sector subcontractors, as well as actors in the ‘migration industry of connectivity services’ (Gordano Peile, 2014), such as money transfer services, mobile phone companies targeting refugees or even illegal organizations exploiting irregular migration. In a context where Western states direct much attention and investment to monitoring and combatting irregular migration in some geographical areas (Andersson, 2016; Słomczyńska & Frankowski, 2016; Triandafyllidou & McAuliffe, 2018), journalistic coverage (BBC, 2021) of the terrifying final hours of a fatal attempt to cross the English channel exemplifies not only the centrality of technology use in life-saving efforts but also the risks digital traces pose to individuals, reflected in them tossing their phone into the waves to protect people traffickers’ identities or to hide details that may prevent their asylum claims being accepted.

The key here is to try to minimize people’s vulnerability in the face of unequal power relations. This may entail very different decisions when trying to better understand labour mobility of highly educated intra-European migrants, or analysing intra-city mobility patterns by vehicle type, rather than when dealing with marginalized and vulnerable groups. Since we know about certain minority populations’ reluctance to participate in routine demographic exercises after experi-

ences of marginalization and stigmatization (Weitzberg, 2015), any choice to make populations visible without informed consent should be carefully considered.

23.3 Concluding Remarks

Based on the above-described potentials and pitfalls of different sources, we recommend that policy-makers use digital data to temper the shortcomings of traditional mobility data – namely, their poor space-time resolution, the limited availability of data disaggregated by sociodemographic characteristics, their delayed availability – when more detailed data is needed to better address policy issues related to inequalities, such as regarding housing, education, health, employment and non-discrimination. Ultimately, the usefulness of digital data and arising methodological challenges depend on research goals. For example, longitudinal mobility estimations using digital data are rendered difficult by changing user bases. No single non-traditional data source captures all types of mobility, but the different sources discussed here capture related but partly different phenomena, including urban and international transport, tourism, population displacements, labour mobility of the highly educated and large-scale mobility data from different data providers, where usability depends on coverage and accessibility. Because of their higher granularity, digital data can monitor and evaluate human mobility and population presence at a higher scale, resolution and detail – in real time – spanning from the micro (local) to the meso (national) and macro (international) level. This is particularly important for policy responses in emergency situations (such as humanitarian or public health emergencies). Here, estimations based on digital data sources can help make faster and more informed decisions. However, in other contexts, it should be carefully considered who benefits if individuals on the move and their practices are made visible.

For models to be correctly specified and for estimations to be reliable, in-depth context-specific knowledge about the ways human mobility occurs on the ground as well as knowledge about quickly changing technology use among the populations of interest is fundamental, although difficult and costly to obtain, and has to be constantly updated. Traditional statistics remain important to evaluate and complement estimations as baselines, especially in a context in which migration is the focus of significant political and media attention and is all too frequently misunderstood or misinterpreted. Moreover, as regards migration, it is crucial that legal policy definitions and normative frameworks are respected. The bottom line is that any model of human movement should be carefully tailored to the specific local context. In an unpredictable world, where people move or not move for myriad reasons and these reasons may vary quickly, as the case of Covid-19 exemplifies, we encourage researchers to constantly reassess if what is being measured reflects the social phenomenon that the measurement is intended to assess and to ensure that their analysis does not generate injustice by rendering people visible in ways that are damaging to their rights and freedoms. This makes the data collection and analysis

process more expensive and less universal than sometimes suggested in relation to new data sources and their usefulness for policy-relevant analysis.

Beyond the realm of migration, digital data on human mobility can assist evidence-based policies on transportation – a primary concern in the field of environmental policies. A well-informed understanding of human mobility and its forms is particularly urgent in a context in which a reduction of fossil fuel-propelled transports in rich countries (flights, cruises and car use) is needed to mitigate global warming (Holden et al., 2019; Peeters & Dubois, 2010). A primary instance is the design of incentives to shift passengers to less polluting travel means – e.g. from airplanes to trains.

Whatever the domain of interest, researchers must be aware that precise, always-on data about individuals, often unobtrusively collected without informed consent, raise several issues concerning privacy and security. Associated risks largely depend on the legal, political and social situations of the individuals or groups eventually covered by it, the actors handling this data and their interests. Operational guidelines on data responsibility are provided, for example, by the Inter-Agency Standing Committee (IASC) of the United Nations system.⁴ Although far from frontline operations, migration research analysing big data can have a quick impact on policies (for instance, border management) and, thus, on human lives – something that traditional studies of migrants rarely had (McAuliffe & Sawyer, 2021). Just because research and policy-making on human mobility have an unprecedented potential to go hand in hand – a good news in itself – we can only urge any actor collecting, using, storing and sharing human mobility data to commit to ‘do no harm while maximizing the benefits’ principles (IASC, 2021), always prioritizing the safe, ethical and effective management of personal and nonpersonal data.

References

- Ahas, R., Silm, S., & Tiru, M. (2018). *Measuring transnational migration with roaming datasets [application/pdf]*. <https://doi.org/10.3929/ETHZ-B-000225599>
- Andersson, R. (2016). Europe’s failed ‘fight’ against irregular migration: Ethnographic notes on a counterproductive industry. *Journal of Ethnic and Migration Studies*, 42(7), 1055–1075. <https://doi.org/10.1080/1369183X.2016.1139446>
- Apple. (2021). COVID-19 – Mobility trends reports. *Apple*. Retrieved from <https://www.apple.com/covid19/mobility>
- Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254–275. <https://doi.org/10.1016/j.trc.2019.02.013>

⁴ They highlight the importance of accountability, confidentiality, coordination and collaboration, data security, necessity and proportionality, fairness and legitimacy, personal data protection, quality, retention and destruction and transparency, within a human rights-based, people-centered and inclusive approach to handling data.

- Bakewell, O. (2008). Research beyond the categories: The importance of policy irrelevant research into forced migration. *Journal of Refugee Studies*, 21(4), 432–453. <https://doi.org/10.1093/jrs/fen042>
- Bao, J., Lian, D., Zhang, F., & Yuan, N. J. (2016). Geo-social media data analytic for user modeling and location-based services. *SIGSPATIAL Special*, 7(3), 11–18. <https://doi.org/10.1145/2876480.2876484>
- Barnett, J., & McMichael, C. (2018). The effects of climate change on the geography and timing of human mobility. *Population and Environment*, 39(4), 339–356. <https://doi.org/10.1007/s11111-018-0295-5>
- Bauman, Z. (1998). *Globalization: The human consequences (Repr.)*. Columbia University Press.
- BBC. (2021, December 21). *Channel migrants tragedy: Terrifying final hours of their fatal journey - BBC News*. News. Retrieved from <https://www.bbc.co.uk/news/resources/idx-b7bd2274-88b1-4ef9-a459-be22e180b52c>
- Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal migration and development: Comparing migration intensities around the world. *Population and Development Review*, 41(1), 33–58. <https://doi.org/10.1111/j.1728-4457.2015.00025.x>
- 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. *PLoS Medicine*, 8(8), e1001083. <https://doi.org/10.1371/journal.pmed.1001083>
- Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1), 1–55. <https://doi.org/10.1140/epjds/s13688-015-0046-0>
- Bloom, L. B. (2020, July 30). Want to live and work in paradise? 7 countries inviting Americans to move abroad. *Forbes*. Retrieved from <https://www.forbes.com/sites/laurabegleybloom/2020/07/30/live-work-remote-move-abroad-coronavirus/>
- Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development*, 18(2), 107–125. <https://doi.org/10.1080/02681102.2011.643209>
- Blumenstock, J., & Fratamico, L. (2013). Social and spatial ethnic segregation: A framework for analyzing segregation with large-scale spatial network data. In *Proceedings of the 4th Annual Symposium on Computing for Development - ACM DEV-4'13*, pp. 1–10. <https://doi.org/10.1145/2537052.2537061>
- Böhme, M. H., Gröger, A., & Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142, 102347. <https://doi.org/10.1016/j.jdeveco.2019.04.002>
- Bojic, I., Sobolevsky, S., Nizetic-Kosovic, I., Podobnik, V., Belyi, A., & Ratti, C. (2016). Sublinear scaling of country attractiveness observed from Flickr dataset. *ArXiv*. Retrieved from <https://dspace.mit.edu/handle/1721.1/109842>
- Bosco, C., Grubanov-Boskovic, S., Iacus, S., Minora, U., Sermi, F., & Spyrtatos, S. (2022). *Data Innovation in Demography, Migration and Human Mobility*. EUR 30907 EN, Publications Office of the European Union: Luxembourg.
- Broeders, D., & Dijstelbloem, H. (2015). The Datafication of mobility and migration management: The mediating state and its consequences. In *Digitizing identities*. Routledge.
- Chareyron, G., Cousin, S., & Jacquot, S. (2021, July 13). L'Europe rythmée par ses visiteurs. 20 ans de commentaires géolocalisés et chronoréférencés. «*Migrer sans entraves*», *De facto [En ligne]*, 27. Retrieved from <https://www.icmigrations.cnrs.fr/2021/06/16/defacto-027-06/>
- Chen, J. J., Mueller, V., Jia, Y., & Tseng, S. K.-H. (2017). Validating migration responses to flooding using satellite and vital registration data. *American Economic Review*, 107(5), 441–445. <https://doi.org/10.1257/aer.p20171052>
- Chi, G., Lin, F., Chi, G., & Blumenstock, J. (2020). A general approach to detecting migration events in digital trace data. *PLoS One*, 15(10), e0239408. <https://doi.org/10.1371/journal.pone.0239408>

- Clemens, M. A., & Ginn, T. (2020). Global mobility and the threat of pandemics: Evidence from three centuries. In *Institute of Labor Economics (IZA), IZA Discussion Papers 13947*.
- Connor, P., & Passel, J. S. (2019). *Europe's unauthorized immigrant population peaks in 2016, then levels off* (p. 53). Pew Research. Retrieved from https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/11/PG_2019.11.13_EU-Unauthorized_FINAL.pdf
- Cui, Y., Meng, C., He, Q., & Gao, J. (2018). Forecasting current and next trip purpose with social media data and Google places. *Transportation Research Part C: Emerging Technologies*, 97, 159–174. <https://doi.org/10.1016/j.trc.2018.10.017>
- de Beer, J., Raymer, J., van der Erf, R., & van Wissen, L. (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe. *European Journal of Population / Revue Européenne de Démographie*, 26(4), 459–481. <https://doi.org/10.1007/s10680-010-9220-z>
- 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. <https://doi.org/10.1038/srep01376>
- Demoli, Y., & Subtil, J. (2019). Boarding Classes. Mesurer la démocratisation du transport aérien en France (1974–2008). *Sociologie*, 10, 2. Retrieved from <https://journals.openedition.org/sociologie/5295>
- EPA. (2019). *Fast facts on transportation greenhouse gas emissions | US EPA*. Retrieved from <https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions>
- European Commission. (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*. Publications Office. Retrieved from <https://data.europa.eu/doi/10.2767/61617>
- Favell, A., & Recchi, E. (2020). Mobilities, neo-nationalism and the lockdown of Europe: Will the European Union survive? In *COMPAS*. Retrieved from <https://www.compas.ox.ac.uk/2020/mobilities-and-the-lockdown-of-europe-will-the-european-union-survive/>
- Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., & Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration. In *Proceedings of the 2017 ACM on Web Science Conference*, pp. 103–110. <https://doi.org/10.1145/3091478.3091496>
- Fiorio, L., Zagheni, E., Abel, G., Hill, J., Pestre, G., Letouzé, E., & Cai, J. (2021). Analyzing the effect of time in migration measurement using georeferenced digital trace data. *Demography*, 58(1), 51–74. <https://doi.org/10.1215/00703370-8917630>
- Florida, R., Rodríguez-Pose, A., & Storper, M. (2021). Cities in a post-COVID world. *Urban Studies*, 004209802110180, 004209802110180. <https://doi.org/10.1177/00420980211018072>
- Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>
- Gabrielli, L., Deutschmann, E., Natale, F., Recchi, E., & Vespe, M. (2019). Dissecting global air traffic data to discern different types and trends of transnational human mobility. *EPJ Data Science*, 8(1), 26. <https://doi.org/10.1140/epjds/s13688-019-0204-x>
- 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. *The VLDB Journal*, 20(5), 695. <https://doi.org/10.1007/s00778-011-0244-8>
- González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782. <https://doi.org/10.1038/nature06958>
- Google. (2021). *COVID-19 community mobility reports*. Retrieved October 03, 2021, from <https://www.google.com/covid19/mobility/>
- Gordano Peile, C. (2014). The migration industry of connectivity services: A critical discourse approach to the Spanish case in a European perspective. *Crossings: Journal of Migration & Culture*, 5(1), 57–71. https://doi.org/10.1386/cjmc.5.1.57_1
- Hannam, K., Sheller, M., & Urry, J. (2006). Editorial: Mobilities, immobilities and moorings. *Mobilities*, 1(1), 1–22. <https://doi.org/10.1080/17450100500489189>

- Hiir, H., Sharma, R., Aasa, A., & Saluveer, E. (2019). Impact of natural and social events on Mobile call data records – An Estonian case study. *Complex Networks and Their Applications VIII Studies in Computational Intelligence*, 882, 415–426. https://doi.org/10.1007/978-3-030-36683-4_34
- Holden, E., Gilpin, G., & Banister, D. (2019). Sustainable mobility at thirty. *Sustainability*, 11(7), 1965. <https://doi.org/10.3390/su11071965>
- Huang, W., Weber, I., & Vieweg, S. (2014). Inferring nationalities of twitter users and studying inter-national linking. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, pp. 237–242. <https://doi.org/10.1145/2631775.2631825>
- Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T., & Wang, F.-Y. (2018). Modeling real-time human mobility based on mobile phone and transportation data fusion. *Transportation Research Part C: Emerging Technologies*, 96, 251–269. <https://doi.org/10.1016/j.trc.2018.09.016>
- Huang, X., Li, Z., Jiang, Y., Ye, X., Deng, C., Zhang, J., & Li, X. (2021). The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. *International Journal of Digital Earth*, 14(4), 424–442. <https://doi.org/10.1080/17538947.2021.1886358>
- Hughes, N. (2018). ‘Tourists go home’: Anti-tourism industry protest in Barcelona. *Social Movement Studies*, 17(4), 471–477. <https://doi.org/10.1080/14742837.2018.1468244>
- Iacus, S. M., Natale, F., Santamaria, C., Spyros, S., & Vespe, M. (2020). Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. *Safety Science*, 129, 104791. <https://doi.org/10.1016/j.ssci.2020.104791>
- IASC (Inter-Agency Standing Committee). (2021). Operational guidance on data responsibility in humanitarian action. <https://interagencystandingcommittee.org/system/files/2021-02/IASC%20Operational%20Guidance%20on%20Data%20Responsibility%20in%20Humanitarian%20Action-%20February%202021.pdf>
- Kikas, R., Dumas, M., & Saabas, A. (2015). Explaining international migration in the Skype network: The role of social network features. In *Proceedings of the 1st ACM Workshop on Social Media World Sensors*, pp. 17–22. <https://doi.org/10.1145/2806655.2806658>
- Kim, J., Sirbu, A., Giannotti, F., & Gabrielli, L. (2020). Digital footprints of international migration on Twitter. In M. R. Berthold, A. Feelders, & G. Kreml (Eds.), *Advances in intelligent data analysis XVIII* (pp. 274–286). Springer. https://doi.org/10.1007/978-3-030-44584-3_22
- Latonero, M., & Kift, P. (2018). On digital passages and Borders: Refugees and the new infrastructure for movement and control. *Social Media + Society*, 4(1), 205630511876443. <https://doi.org/10.1177/2056305118764432>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205.
- Leurs, K., & Smets, K. (2018). Five questions for digital migration studies: Learning from digital connectivity and forced migration in(to) Europe. *Social Media + Society*, 4(1), 2056305118764425. <https://doi.org/10.1177/2056305118764425>
- Li, L., Yang, J., Jing, H., He, Q., Tong, H., & Chen, B. C. (2019). NEMO: Next career move prediction with contextual embedding. In *26th International World Wide Web Conference 2017, WWW 2017 Companion*, pp. 505–513. <https://doi.org/10.1145/3041021.3054200>
- Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A., Canright, G., Engø-Monsen, K., & Bengtsson, L. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change*, 38, 1–7. <https://doi.org/10.1016/j.gloenvcha.2016.02.002>
- Luca, M., Barlacchi, G., Oliver, N., & Lepri, B. (2021). Leveraging mobile phone data for migration flows. *ArXiv:2105.14956 [Cs]*. Retrieved from <http://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. <https://doi.org/10.1007/s41060-017-0065-y>

- McAuliffe, M., & Sawyer, A. (2021). The roles and limitations of data science in understanding international migration flows and human mobility. In *Research handbook on international migration and digital technology*. Retrieved from <https://www.elgaronline.com/view/edcoll/9781839100604/9781839100604.00012.xml>
- Milanović, B. (2019). *Capitalism, alone: The future of the system that rules the world*. The Belknap Press of Harvard University Press.
- Mixed Migration Centre. (2020). *The mixed migration Centre in 2020*. Mixed Migration Centre. Retrieved from https://mixedmigration.org/wp-content/uploads/2021/01/157_annual_catalogue_2020.pdf
- Pachauri, R. K., Allen, M. R., & Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N. K., Edenhofer, O., Elgizouli, I., Field, C. B., Forster, P., Friedlingstein, P., Fuglested, J., Gomez-Echeverri, L., Hallegatte, S., Hegerl, G., Howden, M., et al. (Eds.). (2014). *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the Fifth assessment report of the intergovernmental panel on Climate change*. Intergovernmental Panel on Climate Change.
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 8166. <https://doi.org/10.1038/ncomms9166>
- Peeters, P., & Dubois, G. (2010). Tourism travel under climate change mitigation constraints. *Journal of Transport Geography*, 18(3), 447–457. <https://doi.org/10.1016/j.jtrangeo.2009.09.003>
- Polzer, T., & Hammond, L. (2008). Invisible displacement. *Journal of Refugee Studies*, 21(4), 417–431. <https://doi.org/10.1093/jrs/fen045>
- Pöttschke, S., & Braun, M. (2017). Migrant sampling using Facebook advertisements: A case study of polish migrants in four European countries. *Social Science Computer Review*, 35(5), 633–653. <https://doi.org/10.1177/0894439316666262>
- Rampazzo, F., Bijak, J., Vitali, A., Weber, I., & Zagheni, E. (2021). A framework for estimating migrant stocks using digital traces and survey data: An application in the United Kingdom. *Demography*. Retrieved from <https://eprints.soton.ac.uk/448283/>
- Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, 108(503), 801–819.
- Recchi, E. (2015). *Mobile Europe*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137316028>
- Recchi, E. (2016). Space, mobility and legitimacy. In *Oxford Research Encyclopedia of Politics*. <https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-11?rskey=W5XrDb&result=1>
- Recchi, E., Deutschmann, E., & Vespe, M. (2019). *Estimating transnational human mobility on a global scale*. Robert Schuman Centre for Advanced Studies Research Paper, No 30(RSCAS).
- Recchi, E., Ferrara, A., Rodriguez Sanchez, A., Deutschmann, E., Gabrielli, L., Iacus, S., Bastiani, L., Spyrtatos, S., & Vespe, M. (2022). The impact of air travel on the precocity and severity of COVID-19 deaths in sub-national areas across 45 countries. *Scientific reports*, 12(1), 1–13.
- Rigaud, K. K., de Sherbinin, A., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., Adamo, S., McCusker, B., Heuser, S., & Midgley, A. (2018). *Groundswell: Preparing for internal climate migration* (p. 222). World Bank.
- Scott, J. C. (2008). *Seeing like a state: How certain schemes to improve the human condition have failed (Nachdr.)*. Yale University Press.
- Sirbu, 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. <https://doi.org/10.1007/s41060-020-00213-5>
- Skeldon, R. (2018). *International migration, internal migration, mobility and urbanization: Towards more integrated approaches* (p. 15). *Migration Research Series N° 53*. International Organization for Migration (IOM).

- Słomczyńska, I., & Frankowski, P. (2016). Patrolling power Europe: The role of satellite observation in EU border management. In R. Bossong & H. Carrapico (Eds.), *EU Borders and shifting internal security: Technology, externalization and accountability* (pp. 65–80). Springer. https://doi.org/10.1007/978-3-319-17560-7_4
- Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10), 818–823. <https://doi.org/10.1038/nphys1760>
- Taylor, L. (2023). Data justice, computational social science and policy. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., (Eds.), *Handbook of computational social science for policy*. Springer.
- Tosi, D. (2017). Cell phone big data to compute mobility scenarios for future smart cities. *International Journal of Data Science and Analytics*, 4(4), 265–284. <https://doi.org/10.1007/s41060-017-0061-2>
- Triandafyllidou, A., & McAuliffe, M. (2018). *Migrant smuggling data and research: A global review of the emerging evidence base* (Vol. 2). Retrieved from <https://cadmus.eui.eu/handle/1814/57084>
- Trimikliniotis, N., Parsanoglou, D., & Tsianos, V. (2015). *Mobile commons, migrant digitalities and the right to the city*. Palgrave Macmillan.
- Turton, D. (2005). The meaning of place in a world of movement: Lessons from long-term field research in southern Ethiopia. *Journal of Refugee Studies*, 18(3), 258–280.
- UN DESA. (2015). *International Migration Flows to and from selected countries: The 2015 Revision* [POP/DB/MIG/Flow/Rev.2015]. United Nations, Department of Economic and Social Affairs, Population Division. Retrieved from <https://www.un.org/en/development/desa/population/migration/data/empirical2/docs/migflows2015documentation.pdf>
- UNDESA (2017). *Handbook on Measuring International Migration through Population Censuses*. Retrieved from <https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/international-migration/2017-draft-E.pdf>
- UNDESA. (2019). *International migrant stock 2019*. United Nations. Retrieved from <https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp>
- Urry, J. (2000). Mobile sociology. *The British Journal of Sociology*, 51(1), 185–203. <https://doi.org/10.1111/j.1468-4446.2000.00185.x>
- Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), 565–578. <https://doi.org/10.1002/for.1213>
- Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A.-L. (2011). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1100–1108. <https://doi.org/10.1145/2020408.2020581>
- Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980–991. <https://doi.org/10.1016/j.ijforecast.2014.06.001>
- Weitzberg, K. (2015). The unaccountable census: Colonial enumeration and its implications for the Somali people of Kenya. *The Journal of African History*, 56(3), 409–428. <https://doi.org/10.1017/S002185371500033X>
- Worldbank. (2019). *Mobile cellular subscriptions (per 100 people)—Low & middle income | Data*. Retrieved from https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=XO&most_recent_value_desc=false
- Xiong, C., Hu, S., Yang, M., Younes, H., Luo, W., Ghader, S., & Zhang, L. (2020). Mobile device location data reveal human mobility response to state-level stay-at-home orders during the COVID-19 pandemic in the USA. *Journal of the Royal Society Interface*, 17(173), 20200344. <https://doi.org/10.1098/rsif.2020.0344>
- Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1), 13–25. <https://doi.org/10.1108/IJM-12-2014-0261>
- Zagheni, E., Garimella, V. R. K., Weber, I., & State, B. (2014). Inferring international and internal migration patterns from Twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*, pp. 439–444. <https://doi.org/10.1145/2567948.2576930>

- Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants: Leveraging Facebook's advertising platform. *Population and Development Review*, 43(4), 721–734. <https://doi.org/10.1111/padr.12102>
- Zetter, R. (1991). Labelling refugees: Forming and transforming a bureaucratic identity. *Journal of Refugee Studies*, 4(1), 39–62. <https://doi.org/10.1093/jrs/4.1.39>
- Zumkeller, D. (2009). The dynamics of change: Latest results from the German mobility panel. In *12th International Conference on Travel Behaviour Research*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

