

Chapter 17

Leveraging Digital and Computational Demography for Policy Insights



Ridhi Kashyap and Emilio Zagheni

Abstract Situated at the intersection of the computational and demographic sciences, digital and computational demography explores how new digital data streams and computational methods advance the understanding of population dynamics, along with the impacts of digital technologies on population outcomes, e.g. linked to health, fertility and migration. Encompassing the data, methodological and social impacts of digital technologies, we outline key opportunities provided by digital and computational demography for generating policy insights. Within methodological opportunities, individual-level simulation approaches, such as microsimulation and agent-based modelling, infused with different data, provide tools to create empirically informed synthetic populations that can serve as virtual laboratories to test the impact of different social policies (e.g. fertility policies, support for the elderly or bereaved people). Individual-level simulation approaches allow also to assess policy-relevant questions about the impacts of demographic changes linked to ageing, climate change and migration. Within data opportunities, digital trace data provide a system for early warning with detailed spatial and temporal granularity, which are useful to monitor demographic quantities in real time or for understanding societal responses to demographic change. The demographic perspective highlights the importance of understanding population heterogeneity in the use and impacts of different types of digital technologies, which is crucial towards building more inclusive digital spaces.

R. Kashyap (✉)

Department of Sociology, Leverhulme Centre for Demographic Science, and Nuffield College,
University of Oxford, Oxford, UK
e-mail: ridhi.kashyap@nuffield.ox.ac.uk

E. Zagheni

Max Planck Institute for Demographic Research, Rostock, Germany
e-mail: zagheni@demogr.mpg.de

© 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_17

327

17.1 Introduction

Demography is the scientific study of populations, including the three fundamental forces that shape population dynamics—mortality, fertility and migration. While these three forces produce the essential events for which demographers have developed a range of measurement methods, each of these processes is also the result of complex individual behaviours that are shaped by multiple forces. Thus, in addition to measuring demographic phenomena and describing macrolevel population patterns, demographers examine how and why specific population-level outcomes emerge, seek to explain them and understand their consequences. While pursuing science-driven discovery, demographers also inevitably address several interrelated, policy-relevant themes, such as ageing, family change, the ethnic diversification of societies, spatial segregation and related outcomes and the relationship between environmental and population change. These and related policy-relevant topics are intimately connected with the three core demographic processes of mortality, fertility and migration. For example, significant reductions in mortality rates over the course of the twentieth and twenty-first centuries imply that individuals across Europe can expect to lead long lives, with an increasing overlap of generations within populations. How does the ageing of populations impact on key social institutions linked to the labour market, pension systems and provision of care? Moreover, how can societies better prepare for these changes?

Demography has historically been a data-driven discipline and one that has developed tools to repurpose different kinds of data—often not originally intended or collected for research—for measuring and understanding population change (Billari & Zagheni, 2017; Kashyap, 2021). Demography is thus uniquely positioned to take advantage of the opportunities enabled by the broader development of the computational social sciences, both in terms of new data streams and computational methods. A growing interest in this interface between demography and computational social science has led to the emergence of digital and computational demography (Kashyap et al., 2022). This chapter describes how insights from digital and computational demography can help augment the policy relevance of demographic research.

Demographic research is relevant for policy makers in several ways. At its most basic level, understanding the current as well as anticipated future size, composition and geographical distribution of a population—whether a national, regional or local population—is essential for planning for the provision of services, for identifying targets of aid and for setting policy priorities. For example, the needs for specific public services are closely tied to the age structure of a population—populations that have more young people have very different needs than those with a larger share of older people. The impacts of these age structures are also felt in economic and social domains. This shapes not only what services are needed, e.g. schools versus social care for the elderly, but also which issues require priority at a given time. Demographic analyses can also help identify population changes and trends for the future, to identify areas that will emerge in the future as relevant for policy making.

For instance, subnational areas where population is growing quickly have very different needs, and require a different type of planning, compared to those areas that experience depopulation. More broadly, demography sheds light on population heterogeneity along various dimensions and offers insights into the heterogeneous impact of policy interventions on different segments of the population. For example, when considering key demographic trends like ageing, or the impact of climate change on health and population dynamics, questions of the inequality in these impacts across different regions and socioeconomic groups are critical from a policy perspective for identifying vulnerable communities and for supporting them appropriately. Policy makers may also try to favour certain demographic trends, such as through fertility policies or migration policies, in a way that leads to co-benefits at the individual and societal level. For example, policy makers may pursue fertility policies oriented towards helping individuals achieve the desired number of children, which may in turn affect the long-term sustainability of social security systems.

17.2 The Digital Turn in Demography: An Overview

Demographers have conventionally relied on data sources such as government administrative registers, censuses and nationally representative surveys to describe and understand population trends. A key strength of these data sources that makes them well-suited for demographic research is their representativeness and population generalizability. Censuses and population registers target complete coverage and enumeration of populations. In contrast, the types of surveys conducted by and used by demographers draw on high-quality, probability samples to provide a richer, in-depth source of data with a view to testing specific theories, understanding individual behaviours and attitudes that underpin demographic patterns. While these data sources are critical for demographic research, they also have a number of limitations. These data sources are often slow (e.g. censuses are mostly decennial), resource- and time-intensive and often reactive (e.g. surveys that require asking individuals for information), although in some cases these data are generated as by-products of administrative transactions where individuals interact with state institutions (e.g. birth registration, tax registration). Demographers have developed and applied mathematical and statistical techniques to use quantitative data sources to carefully measure and describe macrolevel (aggregate) population patterns, understand the relationships between different demographic variables and decompose changes in population indicators into different underlying processes. Growing bodies of individual-level and linked datasets have also enabled demographers to address individual-level causal questions about how specific social policies or social changes affect demographic behaviours.

The growing use of digital technologies such as the internet and mobile phones, as well as advances in computational power for processing, storing and analysing

data, has led to a digital and computational turn in demography (Kashyap et al., 2022). This digital turn has affected demographic research along three dimensions:

1. Advancements in data opportunities
2. Applications of computational methods for demographic questions
3. Growing interest in the impacts of digitalization for demographic behaviours

17.2.1 Advances in Data Opportunities

Technological changes in digitized information storage and processing have improved access and granularity of traditional demographic data sources, while also generating new types of data streams and new opportunities for data collection, thereby enriching the demographic data ecosystem (Kashyap, 2021). Some of these new data streams are opened up by the widespread use of digital technologies such as the internet, mobile phones and social media. However, the digitization of information more broadly means that diverse types of digital data sources can now be repurposed for demographic research, ranging from detailed administrative data to bibliometric and crowdsourced genealogical databases, many of which were not intentionally collected for the purpose of research (Alburez-Gutierrez et al., 2019). These new data sources offer novel possibilities, but also come with their own unique ethical and methodological challenges, as we describe in the next section on computational guidelines.

In terms of their opportunities, these new data streams can help fill data gaps in areas where conventional data may be lacking and can provide higher-frequency and real-time measurement than conventional sources of demographic data to capture events as they occur. In addition, they provide better temporal and/or spatial resolution that can help ‘nowcast’ and understand local patterns and indicators in a timely way. For example, a growing body of research has used digital trace data from the web, mobile and social media to measure international or internal migration (e.g. Zagheni & Weber, 2012; Deville et al., 2014; Gabrielli et al., 2019; Alexander et al., 2020; Fiorio et al., 2021; Rampazzo et al., 2021). Different types of digital traces have been used to capture mobility processes. Some widely used examples include aggregated social media audience counts from Facebook’s marketing platform (Rampazzo et al., 2021; Alexander et al., 2020) and timestamped call detail records from mobile phones that provide changing spatiotemporal distributions of mobile users (e.g. Deville et al., 2014). Vehicle detection with machine learning (ML) techniques applied to satellite images obtained via remote sensing have also been used to track mobility processes (e.g. Chen et al., 2014). Conventional data on migration are often lacking, and these studies identify ways in which these nontraditional data can help fill gaps and complement traditional sources of demographic statistics. Digital traces of behaviours, such as those from aggregate web search queries or social media posts, can further provide non-elicited forms of measurement of contexts, norms and behaviours that are relevant for understanding demographic shifts (Kashyap, 2021). For example, aggregated web search queries

have been shown to capture fertility intentions that are predictive of fertility rates (Billari et al., 2016; Wilde et al., 2020) or information-seeking about abortion (Reis & Brownstein, 2010; Leone et al., 2021). Social media posts have also been used to study sentiments surrounding parenthood (Mencarini et al., 2019), while satellite images have been used to assess the socioeconomic characteristics of geographical areas (Elvidge et al., 2009; Gebru et al., 2017; Jochem et al., 2021).

Beyond passive measurement from already existing digital traces, internet- and mobile-based technologies can also provide cost-effective opportunities for data collection. Targeted recruitment of survey respondents, based on social and demographic attributes such as those provided by the social media advertisement platforms (e.g. Facebook), has enabled research on hard-to-reach groups, e.g. migrant populations (Pötzschke & Braun, 2017), or those working in specific service sector jobs/occupations (Schneider & Harknett, 2019a, b). Digital modes of data collection also proved invaluable during the COVID-19 pandemic, when rapid understanding of social and behavioural responses to the pandemic and associated lockdowns was needed but traditional face-to-face forms of data collection were impossible (Grow et al., 2020). Combining passively collected information (e.g. from social media or mobile phones) with accurate surveys is an active area of research, with great promise in the context of monitoring indicators of sustainable development on a global scale (Kashyap et al., 2020; Aiken et al., 2022; Chi et al., 2022).

17.2.2 Computational Methods for Demographic Questions

Second, improvements in computational power have facilitated the adoption of computational methodologies, such as microsimulation and agent-based simulation, as well as ML techniques, for demographic applications. Microsimulation techniques, which take empirical transition rates of mortality, fertility and migration as their input to generate a synthetic population that has a realistic genealogical structure, have been used to study the evolution of population dynamics. Microsimulation techniques have been used to examine kinship dynamics and intergenerational processes, such as the availability and potential support of kin and extended family across the life course (Zagheni, 2010; Verdery & Margolis, 2017; Verdery, 2015) or the extent of generational overlap (Alburez-Gutierrez et al., 2021), as well as the impact of macrolevel changes, like technological changes (Kashyap & Villavicencio, 2016) or educational change (Potančoková & Marois, 2020) that affect demographic rates, on population dynamics. Agent-based simulation techniques build on microsimulation by incorporating individual-level behavioural rules, social interaction and feedback mechanisms to test behavioural theories for how macrolevel population phenomena emerge from individual-level behaviours. Agent-based simulation approaches have been used within the demographic literature to model migration decision-making (Klabunde & Willekens, 2016; Entwisle et al., 2016) as well as family and marriage formation processes (Billari et al., 2007;

Diaz et al., 2011; Grow & Van Bavel, 2015). Both these types of individual-based simulation techniques that model individual-level probabilities of experiencing events—when infused with different types of real demographic data—offer ways of building what Bijak et al. describe as ‘semi-artificial’ population models that are empirically informed (Bijak et al., 2013). Such semi-artificial models are useful for generating scenarios to examine social interaction and feedback effects or assess the likely consequences of policies given a set of theoretical expectations. These approaches can be used to generate synthetic counterfactual scenarios and are useful to identify causal relationships, especially for social and demographic questions for which experimental approaches like randomized trials are not possible nor ethically desirable. Given that policy making is often concerned with causal relationships, these tools are of quintessential importance, especially when used in combination with data-driven approaches.

Improvements in computational power combined with an increasingly data-rich environment have also opened up opportunities for the use of ML approaches in demographic research. The focus on the discovery of macrolevel regularities in population dynamics, its interest in exploring different dimensions of population heterogeneity and the discipline’s orientation towards projection of unseen (future) trends based on seen (past) trends lends itself well to the applications of ML techniques (Kashyap et al., 2022). An emerging body of work has used supervised ML approaches that find predictive models that link some explanatory variables to some outcome to individual-level longitudinal survey data to assess the predictability of demographic and life course outcomes (Salganik et al., 2020; Arpino et al., 2022). ML approaches have also been used for demographic forecasting (Nigri et al., 2019; Levantesi et al., 2022) and for population estimation using geospatial data (Stevens et al., 2015; Lloyd et al., 2017). While demographic research has been broadly concerned with prediction of risk for population groups or sub-groups, ML techniques offer the opportunity to generate more accurate predictions at the individual level and to better quantify heterogeneity in outcomes or responses.

17.2.3 Demographic Impacts of Digitalization

Digitalization has implications for demographic processes as digital tools are used for information-seeking, social interaction and communication and accessing vital services. The importance of digital technologies as a lifeline for different domains was powerfully illuminated during the COVID-19 pandemic. Demographic research has highlighted how the use of internet and mobile technologies can directly impact on demographic outcomes linked to health (Rotondi et al., 2020), marriage (Bellou, 2015; Sironi & Kashyap, 2021), fertility (Billari et al., 2019, 2020) and migration (Pesando et al., 2021), by enabling access to information, promoting new paths for social learning and interaction and providing flexibility in reconciling work and family (e.g. through remote working). This research suggests that access to digital resources (e.g. broadband connectivity, mobile apps) may, for example,

enhance the health, wellbeing and quality of life in sparsely populated areas, by enabling better connectivity, access to services and economic opportunities in those regions. This may contribute to reduce depopulation in certain rural areas of Europe, by making them more attractive places to live and work. At the same time, not everyone may have the same level of access or skills necessary to take full advantage of the digital revolution (van Deursen & van Dijk, 2011; Alvarez-Galvez et al., 2020), and a deeper examination of the heterogeneity of these impacts is necessary to understand who and under what conditions digital technologies can empower. In addition to understanding the social impacts of digital technologies, there is value in understanding the demographic characteristics of digital divides also from the perspective of using new streams of digital data for population generalizable measurement. This is an area where demographers have also begun to make contributions through exploring demographic dimensions of social media and internet use (Feehan & Cobb, 2019; Gil-Clavel & Zagheni, 2019; Kashyap et al., 2020).

17.3 Computational Guidelines

Digital and computational demography, which bridges computational social science with demography, offers several opportunities for addressing policy-relevant questions. We provide guidelines for leveraging these opportunities along three dimensions: methodological opportunities, data opportunities and understanding demographic heterogeneity in the impacts of digital technologies.

17.3.1 *Methodological Opportunities*

Policy makers frequently need to understand the impacts of specific policies or a basket of policies (e.g. fertility policies that seek to promote the realization of desired fertility), examine multiple scenarios and counterfactuals and assess the heterogeneity in the impacts of specific policies or social and environmental changes (e.g. climate change) on populations. Computational simulation techniques such as microsimulation and agent-based simulation, which have been increasingly adopted within digital and computational demography, are particularly useful for addressing these types of questions. By incorporating different types of data and forms of population heterogeneity (e.g. differences by educational groups) within simulation models, these approaches can be used to create synthetic populations where individual decisions and behaviours are guided by empirical survey data and/or observed demographic rates (e.g. birth, death or migration rates).

Agent-based simulation approaches are especially useful when the focus is on understanding non-linear feedback effects or social influence effects on behaviours, such as those linked to whether or not to have a child given a wider set of

contextual conditions. Microsimulation approaches can help understand the broader implications of a current set of demographic rates for population composition and change, as well as for kinship and intergenerational processes. Microsimulation techniques, for example, can help understand the evolution of kinship availability and support as a consequence of changing demographic rates. By incorporating rates that vary by different population sub-groups (e.g. ethnic groups), microsimulation approaches can help explore questions about the future size and composition of the availability of kin support for different population groups, which is a central question for understanding and adapting in the context of population ageing. These approaches provide the necessary flexibility to create counterfactual scenarios and for an opportunity to link different types of data to understand how different parts of a population system respond—e.g. individual-level changes affect macrolevel patterns, or macrolevel shocks affect individuals. A central challenge when building simulation models is the trade-off between parsimony and complexity. On the one hand, while simulation models allow for flexibility to incorporate different parameters to model complex systems, the inclusion of too many parameters can be counterproductive for interpretability, i.e. for understanding which parameters directly affect the outcome of interest. Another separate concern is that of how best to understand model uncertainty and draw statistical inferences from model outcomes. To this end, different approaches for computationally intensive calibration of simulation models have been applied within the demographic literature. These approaches combine the tools of statistics (including Bayesian statistics) with simulation approaches to help assess model sensitivity and uncertainty (Poole & Raftery, 2000; Bijak et al., 2013).

As noted in the previous section, the data ecosystem of demography has been significantly enriched with the digital revolution. The availability of a greater variety of data sources and the ability to link them, either at individual or aggregate levels, offer an opportunity to apply tools of causal analysis for observational data, such as quasi-experimental techniques. These techniques can be especially powerful for analysing the impacts of climate shocks (e.g. temperature changes, natural disasters). Such research designs are enabled by the availability of georeferenced data and the ability to link these to other data, e.g. survey or census datasets, thereby facilitating analysis of the impacts of environmental contexts on demographic outcomes (e.g. Andriano & Behrman, 2020; Hauer et al., 2020; Thiede et al., 2022).

Computational methods like ML further provide new approaches to harness an enriched data ecosystem. While a lot of social demographic research has been guided by a theoretical perspective focused on analysing the specific relationship between a theoretical predictor and outcome of interest, ML techniques allow for ways in which a wider range of potential predictors (or features) that are increasingly available in our data sources as well as different functional forms can inform analyses such that new patterns can be learned from data. From a policy perspective, these approaches have the potential to help identify new types of regularities and relationships between variables (e.g. social factors and health outcomes), detect vulnerable population sub-groups and help guide new questions to identify new social mechanisms that can help streamline the targeting and delivery of public

services and social policies (e.g. Wang et al., 2013; Mhasawade et al., 2021; Aiken et al., 2022). The deployment of algorithmic decision-making processes however also raises significant social and ethical challenges, such as those about bias and discrimination, whereby algorithms can amplify existing patterns of social disadvantage, as well as transparency and accountability, particularly given concerns about the opacity of complex algorithms (Lepri et al., 2018). Insights from the demographic literature further emphasize the importance of proceeding carefully when deploying these tools. Social demographic research that has applied ML techniques to long-standing survey datasets to predict life course outcomes such as educational performance or material hardship has shown that these outcomes are often challenging to predict at the individual level (Salganik et al., 2020). More work is needed to understand the conditions under which ML approaches can help improve predictive accuracy with different types of social data but also to better evaluate the social and ethical implications and trade-offs in the use of predictive approaches for policy making.

17.3.2 Data Opportunities

Policy makers are interested in knowing about real-time developments as they unfold. A key challenge with traditional sources of demographic data, as noted in the previous section, has often been their slower timeliness and lags between data collection, processing and publication. Digital trace data, which are generated as by-products of the use of web, social media and mobile technologies, are often able to more effectively capture real-time processes. The widespread use of different types of digital technologies in different domains of life implies that aggregated forms of these data can provide meaningful signals of population behaviours. For example, the reliance on search engines such as Google for information-seeking means that aggregated web search queries, such as those provided via Google Trends, can help us understand health concerns or behaviours, or fertility intentions within a population. When calibrated to ‘ground truth’ demographic data sources, these real-time data have the potential to help predict future changes and ‘nowcast’ patterns before they appear in official statistics.

More generally, new data opportunities provide a system for early warning with detailed spatial and temporal granularity. This can be useful in cases where demographic quantities, like migration flows, need to be monitored in response to a crisis, or for understanding the societal responses to demographic change, e.g. misinformation related to migration or media portrayals of immigrant populations. The value of nontraditional, digital trace datasets for monitoring mobility was highlighted during the COVID-19 pandemic, where Google mobility data was used to track the impacts of lockdowns and for other forms of public health surveillance (Google, 2022). These data proved useful to assess the potential impact of policy decisions related to partial or full lockdowns, and related reductions in mobility, on lives saved (Basellini et al., 2021).

Different types of digital trace data can also provide complementary measures of sentiments, attitudes, norms and current conversations in different formats (e.g. images, text) that are useful for capturing social responses to events, as might be required for policy makers. Online spaces have become salient spaces for social interaction and exchange, information-seeking and collective expression and mobilization. For example, in the area of fertility and family formation, online platforms and forums, such as Mumsnet or fertility apps, can provide a view on prevailing sentiments, concerns and aspirations surrounding parenthood. For other domains, such as for the labour market, when understanding supply or demand in specific sectors may be necessary (e.g. long-term care), online job search forums can provide insights into these dynamics (e.g. Buchmann et al., 2022). Social media can also provide a useful barometer to track sentiments surrounding immigration or policy changes surrounding immigration (e.g. Flores, 2017) while also providing novel ways to measure the integration of immigrant groups (e.g. Dubois et al., 2018).

While digital trace data provide unique opportunities, it is important to ensure that appropriate ethical, measurement and theoretical frameworks guide the use of the data for policy purposes and, where feasible, the data be triangulated and contextualized against traditional data sources. In many cases, aggregated data are sufficient to address a policy-relevant research question, whereas in others more fine-grained, individual-level information may be needed. In cases where aggregated data are insufficient, creating ways to appropriately anonymize the data and safeguard against any risk of harming respondents should remain priority. Given that digital trace data are often not expressly collected for research and collected with informed consent, which is a fundamental principle for survey research, higher standards of privacy protection should be adhered to when using these data. A central challenge with digital trace data remains data access. These data come from and are often owned by private companies, which implies that both their access can often be limited and important details of the proprietary algorithms that shape them may not be known. The landscape of access to digital trace data, via more democratic modes of access such as public application programming interfaces (APIs), has become increasingly more constrained, and in many cases platform terms of use have become more stringent. Policy initiatives to support the development of transparent frameworks for enabling ethically guided and privacy-preserving modes of data sharing between research institutions and private companies are urgently needed to ensure that the potential of these data is realized.

When analysing digital traces, it is important to consider demographic biases to better understand who is represented in them and the broader generalizability of the data. These biases may reflect broader digital divides in internet access or platform-specific patterns of use. Triangulation against high-quality traditional data, e.g. from probability surveys, can be valuable in assessing these biases. A separate, but equally important, consideration is that of algorithmic bias, i.e. whereby algorithms implemented on online platforms shape behaviours, such that it is difficult to assess whether observed patterns detected in the data reflect actual behaviours or the algorithms. One way to address algorithmic bias is to move beyond passively

collected digital traces towards data collection that involves surveying respondents directly, as we describe next.

The increasing adoption of digital technologies has also facilitated online and mobile modes for primary data collection. For example, even in the case of traditional data sources such as censuses, respondents can fill in questionnaires online, although no census so far has shifted completely online as the exclusive mode of data collection. Digital technologies provide cost-efficient modes for survey data collection, although mode and demographic biases of these platforms need to be addressed when using these approaches. A significant opportunity for online recruitment of specific population groups, e.g. migrants or new parents, is provided by social media-targeted advertisement platforms. These are relevant from a policy perspective as they offer new opportunities for data collection that are cost-efficient, timely, and can help overcome some of the limitations of only passively collected digital traces. For example, Facebook allows ads that are targeted towards migrants from specific countries or language speakers, although the algorithms used to determine whether a user is a migrant are unclear. By conducting surveys on migrant groups where respondents are recruited using these algorithmic targeting capabilities of social media ad platforms, researchers can help audit the algorithms that are used in designing the targeting features of these platforms. While such online surveys offer advantages, they are not high-quality probability samples. Drawing population-level inferences from them requires users to collect demographic information within them followed by the application of de-biasing techniques such as post-stratification weighting, where population weights come from a source such as a census or a high-quality probability survey (Zagheni & Weber, 2015).

An important direction for extracting greater value from digital behavioural data is to integrate these with surveys—for example, mobile app-based modes of data collection may enable both the collection of self-reported information combined with data on location or movement (e.g. via an accelerometer) or time use. More broadly, data linkage of different types of data—e.g. survey and geospatial data, administrative data with survey data—can help bolster the value that can be derived from data for policy purposes. Linked administrative data, such as that from population registers, are a key resource for demographic research. The Nordic countries (Thomsen & Holmøy, 1998; Blom & Carlsson, 1999), but also others such as the Netherlands (Bakker et al., 2014), have led the way in creating robust data infrastructures and access to these data, and greater policy efforts across Europe to improve linkage of and access to administrative data are highly desirable.

17.3.3 Understanding Demographic Heterogeneity in the Impacts of Digital Technologies

Research suggests that digital technologies, by providing cost-effective ways of accessing information, enabling communication and exchange and providing access to vital services, can help empower individuals in different domains of life, including their health, wellbeing and family life, among others. Digital technologies have the potential to provide valuable tools, for example, for mitigating isolation and exclusion of rural or ageing populations, or providing modes for flexible working. While technology has the potential to make significant positive impacts, the internet is also not a singular technology, and one where content is often deregulated and user-generated and where the risk of misinformation is also present. From a policy perspective to ensure that the full potential of digital technologies is realized effectively and equitably, it is essential to understand who is using digital technologies and tools (or not), how they use them and who benefits from them. The demographic perspective can be especially valuable for understanding this with the aim of clarifying who and under what conditions technology can empower and when it does not.

For understanding demographic differences in the use of digital technologies and functionalities, different data sources are needed. First, a deeper assessment of these differences requires more detailed questions, moving beyond simple measures of internet use within traditional data sources, e.g. large-scale social and demographic survey data infrastructures, to understand how individuals are leveraging technologies for various life domains. Second, administrative data from governments, but also private companies (e.g. mobile phone operators), can provide important insights on the use of digital services by demographic groups. Policy makers should seek to incorporate demographic information (e.g. age, gender, education, ethnicity) where possible in identifying the uptake and impacts of digital tools. Third, digital traces from different platforms can themselves be useful for understanding demographic differences in the use of different platforms in some cases. For example, data from the social media marketing platforms can provide insights on the demographic composition of their user base, although the aforementioned limitations about potential algorithmic bias affecting these data should be carefully considered when interpreting these data.

17.4 Discussion

Demography is a highly policy-relevant discipline. As this chapter has highlighted, the new data sources and computational tools available to demographers enable us to provide sharper images of our societies and of sociodemographic mechanisms. This, in turn, amplifies our intuition of the implications of alternative policy choices. While the use of computational approaches, such as those outlined in this

chapter, is clearly valuable, we emphasize that these are best thought as providing complementary and synergistic potential. The most fruitful use cases are likely to be those where both traditional and nontraditional data can be integrated for policy making purposes.

Computational modelling approaches that we have described, such as individual-level simulation models, will further benefit from integrating different types of data to help build ‘semi-artificial’ societies (Bijak et al., 2013), or in other words empirically informed synthetic models, that can serve as virtual laboratories to assess the potential social impacts of different policies. These provide useful tools to assess policy-relevant questions about the impacts of the future course of key demographic trends, such as ageing, climate change and immigration.

A distinct opportunity offered by the demographic perspective is the importance of understanding demographic differences in the use of different types of digital technologies and platforms. This is crucial both from the perspective of understanding their social impacts and also for more careful use, analysis and interpretation of the data generated by the use of technologies (e.g. digital trace data). The internet is not a singular technology, yet the digital revolution has affected nearly all domains of life. Understanding population-level heterogeneity in digital access and skills, as well as identifying pathways through which digital tools can empower different marginalized populations (e.g. rural populations, older populations), is crucial for addressing population inequalities. Ensuring that no one is left behind in digital spaces is something that needs to be addressed by policy makers, as presently significant digital divides in digital infrastructure, as well as digital skills, persist, such as between Eastern and Western Europe (OECD, 2019). Closing these divides will require policy efforts targeting both infrastructure and also digital (up)-skilling to facilitate the digital inclusion of communities.

Policy efforts that push for frameworks for data sharing and access between researchers and proprietary datasets to facilitate their scientific use are crucial for realizing the opportunities offered by new types of data. The involvement of researchers, not only at point of access but also in the process of coproduction of proprietary datasets and for algorithmic transparency, is desirable, to ensure constructive use for scientific and policy insights. Beyond proprietary data, the data revolution also encompasses administrative data held by governments, which is now increasingly digitized, and streamlined access to these data as well as frameworks to facilitate more effective data linkage between different governmental agencies is crucial. While the data ecosystem has diversified and become enriched, we stress that more and bigger datasets do not necessarily mean better data. The proper assessment of data quality and reliance on proper measurement should remain core principles when collecting, producing, using and analysing data, which are areas where demographic research has much to contribute. Lastly, it is useful to remember that while better data when used in an ethical way can provide better images of our societies, data itself can only help us identify problems, but does not solve them.

References

- Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, *603*(7903), 864–870. <https://doi.org/10.1038/s41586-022-04484-9>
- Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). Demography in the digital era: New data sources for population research. Preprint. *SocArXiv*:<https://doi.org/10.31235/osf.io/24jp7>.
- Alburez-Gutierrez, D., Mason, C., & Zagheni, E. 2021. The “sandwich generation” revisited: Global demographic drivers of care time demands. *Population and Development Review*. Advanced Publication. doi:<https://doi.org/10.1111/padr.12436>.
- Alexander, M., Polimis, K., & Zagheni, E. (2020). Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review*, August. doi:<https://doi.org/10.1007/s11113-020-09599-3>.
- Alvarez-Galvez, J., Salinas-Perez, J. A., Montagni, I., & Salvador-Carulla, L. (2020). The persistence of digital divides in the use of health information: A comparative study in 28 European countries. *International Journal of Public Health*, *65*(3), 325–333. <https://doi.org/10.1007/s00038-020-01363-w>
- Andriano, L., & Behrman, J. (2020). The effects of growing-season drought on young women’s life course transitions in a Sub-Saharan context. *Population Studies*, *74*(3), 331–350. <https://doi.org/10.1080/00324728.2020.1819551>
- Arpino, B., Le Moglie, M., & Mencarini, L. (2022). What tears couples apart: A machine learning analysis of union dissolution in Germany. *Demography*, *59*(1), 161–186. <https://doi.org/10.1215/00703370-9648346>
- Bakker, B. F. M., van Rooijen, J., & van Toor, L. (2014). The system of social statistical datasets of statistics Netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the IAOS*, *30*(4), 411–424. <https://doi.org/10.3233/SJI-140803>
- Basellini, U., Alburez-Gutierrez, D., Del Fava, E., Perrotta, D., Bonetti, M., Camarda, C. G., & Zagheni, E. (2021). Linking excess mortality to mobility data during the first wave of COVID-19 in England and Wales. *SSM - Population Health*, *14*, 100799. <https://doi.org/10.1016/j.ssmph.2021.100799>
- Bellou, A. (2015). The impact of internet diffusion on marriage rates: Evidence from the broadband market. *Journal of Population Economics*, *28*(2), 265–297. <https://doi.org/10.1007/s00148-014-0527-7>
- Bijak, J., Hilton, J., Silverman, E., & Cao, V. D. (2013). Reforging the wedding ring: Exploring a semi-artificial model of population for the United Kingdom with Gaussian process. *Demographic Research*, *29*, 729–766. <https://doi.org/10.4054/DemRes.2013.29.27>
- Billari, F. C., & Zagheni, E. (2017). *Big data and population processes: A revolution?*, July. doi:<https://doi.org/10.31235/osf.io/f9vzp>.
- Billari, F. C., Prskawetz, A., Diaz, B. A., & Fent, T. (2007). The “wedding-ring”: An agent-based marriage model based on social interaction. *Demographic Research*, *17*, 59–82.
- Billari, F. C., D’Amuri, F., & Marcucci, J. (2016). Forecasting births using Google. In *CARMA 2016: 1st International Conference on Advanced Research Methods in Analytics* (pp. 119–119). Editorial Universitat Politècnica de València. doi:<https://doi.org/10.4995/CARMA2016.2015.4301>.
- Billari, F. C., Giuntella, O., & Stella, L. (2019). Does broadband internet affect fertility? *Population Studies*, *73*(3), 297–316. <https://doi.org/10.1080/00324728.2019.1584327>
- Billari, F. C., Rotondi, V., & Trinitapoli, J. (2020). Mobile phones, digital inequality, and fertility: Longitudinal evidence from Malawi. *Demographic Research*, *42*, 1057–1096.
- Blom, E., & Carlsson, F. (1999). Integration of administrative registers in a statistical system: A Swedish perspective. *Statistical Journal of the United Nations Economic Commission for Europe*, *16*(2–3), 181–196. <https://doi.org/10.3233/SJU-1999-162-307>

- Buchmann, M., Buchs, H., Busch, F., Clematide, S., Gnehm, A.-S., & Müller, J. (2022). Swiss job market monitor: A rich source of demand-side micro data of the labour market. *European Sociological Review*, January, jcac002. doi: <https://doi.org/10.1093/esr/jcac002>.
- Chen, X., Xiang, S., Liu, C.-L., & Pan, C.-H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1797–1801. <https://doi.org/10.1109/LGRS.2014.2309695>
- Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3), e2113658119. <https://doi.org/10.1073/pnas.2113658119>
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888–15893. <https://doi.org/10.1073/pnas.1408439111>
- Diaz, B. A., Fent, T., Prskawetz, A., & Bernardi, L. (2011). Transition to parenthood: The role of social interaction and endogenous networks. *Demography*, 48(2), 559–579. <https://doi.org/10.1007/s13524-011-0023-6>
- Dubois, A., Zagheni, E., Garimella, K., & Weber, I. (2018). Studying migrant assimilation through Facebook interests. In S. Staab, O. Koltsova, & D. I. Ignatov (Eds.), *Social Informatics* (pp. 51–60). Lecture Notes in Computer Science. Springer International Publishing. doi:https://doi.org/10.1007/978-3-030-01159-8_5.
- Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers and Geosciences*, 35(8), 1652–1660. <https://doi.org/10.1016/j.cageo.2009.01.009>
- Entwisle, B., Williams, N. E., Verdery, A. M., Rindfuss, R. R., Walsh, S. J., Malanson, G. P., Mucha, P. J., et al. (2016). Climate shocks and migration: An agent-based modeling approach. *Population and Environment*, 38(1), 47–71. <https://doi.org/10.1007/s11111-016-0254-y>
- Feehan, D. M., & Cobb, C. (2019). Using an online sample to estimate the size of an offline population. *Demography*, 56(6), 2377–2392. <https://doi.org/10.1007/s13524-019-00840-z>
- 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>
- Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter data. *American Journal of Sociology*, 123(2), 333–384. <https://doi.org/10.1086/692983>
- 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>
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 114(50), 13108–13113.
- Gil-Clavel, S., & Zagheni, E. (2019). Demographic differentials in Facebook usage around the world. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(July), 647–650.
- Google. (2022). *COVID-19 community mobility report*. 2022. <https://www.google.com/covid19/mobility?hl=en>
- Grow, A., & Van Bavel, J. (2015). Assortative mating and the reversal of gender inequality in education in Europe: An agent-based model. Edited by Hemachandra Reddy. *PLoS One*, 10(6), e0127806. <https://doi.org/10.1371/journal.pone.0127806>
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). Addressing public health emergencies via Facebook surveys: Advantages, challenges, and practical considerations. *Journal of Medical Internet Research*, 22(12), e20653. <https://doi.org/10.2196/20653>
- Hauer, M. E., Holloway, S. R., & Oda, T. (2020). Evacuees and migrants exhibit different migration systems after the Great East Japan earthquake and tsunami. *Demography*, 57(4), 1437–1457. <https://doi.org/10.1007/s13524-020-00883-7>

- Jochem, W. C., Leasure, D. R., Pannell, O., Chamberlain, H. R., Jones, P., & Tatem, A. J. (2021). Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science*, 48(5), 1161–1179. <https://doi.org/10.1177/2399808320921208>
- Kashyap, R. (2021). Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. *Population Studies*, 75(sup1), 47–75. <https://doi.org/10.1080/00324728.2021.1969031>
- Kashyap, R., & Villavicencio, F. (2016). The dynamics of son preference, technology diffusion, and fertility decline underlying distorted sex ratios at birth: A simulation approach. *Demography*, 53(5), 1261–1281. <https://doi.org/10.1007/s13524-016-0500-z>
- Kashyap, R., Fatehikia, M., Al Tamime, R., & Weber, I. (2020). Monitoring global digital gender inequality using the online populations of Facebook and Google. *Demographic Research*, 43, 779–816.
- Kashyap, R., Gordon Rinderknecht, R., Akbaritabar, A., Alburez-Gutierrez, D., Gil-Clavel, S., Grow, A., Kim, J., et al. (2022). Digital and computational demography. *SocArXiv*. <https://doi.org/10.31235/osf.io/7bvpt>
- Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, 32(1), 73–97. <https://doi.org/10.1007/s10680-015-9362-0>
- Leone, T., Coast, E., Correa, S., & Wenham, C. (2021). Web-based searching for abortion information during health emergencies: A case study of Brazil during the 2015/2016 Zika Outbreak. *Sexual and Reproductive Health Matters*, 29(1), 1883804. <https://doi.org/10.1080/26410397.2021.1883804>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy and Technology*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>
- Levantesi, S., Nigri, A., & Piscopo, G. (2022). Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. *International Journal of Approximate Reasoning*, 140(January), 282–297. <https://doi.org/10.1016/j.ijar.2021.10.008>
- Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific Data*, 4(1), 1–17. <https://doi.org/10.1038/sdata.2017.1>
- Mencarini, L., Hernández-Farías, D. I., Lai, M., Patti, V., Sulis, E., & Vignoli, D. (2019). Happy parents' tweets: An exploration of Italian twitter data using sentiment analysis. *Demographic Research*, 40, 693–724.
- Mhasawade, V., Zhao, Y., & Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8), 659–666. <https://doi.org/10.1038/s42256-021-00373-4>
- Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., & Perla, F. (2019). A deep learning integrated Lee–Carter model. *Risks*, 7(1), 33. <https://doi.org/10.3390/risks7010033>
- OECD. (2019). Skills for a digital society. In *OECD skills outlook 2019: Thriving in a digital world*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/oecd-skills-outlook-2019_df80bc12-en.
- Pesando, L. M., Rotondi, V., Stranges, M., Kashyap, R., & Billari, F. C. (2021). The internetization of international migration. *Population and Development Review*, 47(1), 79–111. <https://doi.org/10.1111/padr.12371>
- Poole, D., & Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, 95(452), 1244–1255.
- Potaňčoková, M., & Marois, G. (2020). Projecting future births with fertility differentials reflecting women's educational and migrant characteristics. *Vienna Yearbook of Population Research*, 18, 141–166.
- 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*.
- Reis, B. Y., & Brownstein, J. S. (2010). Measuring the impact of health policies using internet search patterns: The case of abortion. *BMC Public Health*, 10(1), 514. <https://doi.org/10.1186/1471-2458-10-514>
- Rotondi, V., Kashyap, R., Pesando, L. M., Spinelli, S., & Billari, F. C. (2020). Leveraging mobile phones to attain sustainable development. *Proceedings of the National Academy of Sciences*, 117(24), 13413–13420. <https://doi.org/10.1073/pnas.1909326117>
- Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), 8398–8403. <https://doi.org/10.1073/pnas.1915006117>
- Schneider, D., & Harknett, K. (2019a). Consequences of routine work-schedule instability for worker health and well-being. *American Sociological Review*, 84(1), 82–114. <https://doi.org/10.1177/0003122418823184>
- Schneider, D., & Harknett, K. (2019b). What's to like? Facebook as a tool for survey data collection. *Sociological Methods and Research*, November, 0049124119882477. doi:<https://doi.org/10.1177/0049124119882477>.
- Sironi, M., & Kashyap, R. (2021). Internet access and partnership formation in the United States. *Population Studies*, November, 1–19. <https://doi.org/10.1080/00324728.2021.1999485>
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, 10(2), e0107042. <https://doi.org/10.1371/journal.pone.0107042>
- Thiede, B. C., Randell, H., & Gray, C. (2022). The childhood origins of climate-induced mobility and immobility. *Population and Development Review*. <https://doi.org/10.1111/padr.12482>
- Thomsen, I., & Holmøy, A. M. K. (1998). Combining data from surveys and administrative record systems. The Norwegian experience. *International Statistical Review*, 66(2), 201–221. <https://doi.org/10.1111/j.1751-5823.1998.tb00414.x>
- van Deursen, A., & van Dijk, J. (2011). Internet skills and the digital divide. *New Media and Society*, 13(6), 893–911. <https://doi.org/10.1177/1461444810386774>
- Verdery, A. M. (2015). Links between demographic and kinship transitions. *Population and Development Review*, 41(3), 465–484. <https://doi.org/10.1111/j.1728-4457.2015.00068.x>
- Verdery, A. M., & Margolis, R. (2017). Projections of white and black older adults without living kin in the United States, 2015 to 2060. *Proceedings of the National Academy of Sciences*, 114(42), 11109–11114. <https://doi.org/10.1073/pnas.1710341114>
- Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Learning to detect patterns of crime. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine learning and knowledge discovery in databases* (pp. 515–530). Lecture Notes in Computer Science. Springer. doi:https://doi.org/10.1007/978-3-642-40994-3_33.
- Wilde, J., Chen, W., & Lohmann, S. (2020). COVID-19 and the future of US fertility: What can we learn from Google? *Working Paper 13776. IZA Discussion Papers*. <https://www.econstor.eu/handle/10419/227303>.
- Zagheni, E. (2010). *The impact of the HIV/AIDS epidemic on orphanhood probabilities and kinship structure in Zimbabwe*. UC Berkeley. <https://portal.demogr.mpg.de/uc/item/,DanaInfo=escholarship.org,SSL+7xp9m970>.
- Zagheni, E., & Weber, I. (2012). You are where you E-Mail: Using e-Mail data to estimate international migration rates. In *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 348–351). WebSci '12. Association for Computing Machinery. doi:<https://doi.org/10.1145/2380718.2380764>.
- Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, 36(1), 13–25.

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.

