UCLPRESS

Article title: Key opportunities and challenges for the use of big data in migration research and policy
Authors: Lydia Franklinos[1], Rebecca Parrish[2], Rachel Burns[3], Andrea Caflisch[4], Bishawjit Mallick[5], Taifur Rahman[6], Vasileios Routsis[7], Ana Sebastián López[8], Andrew Tatem[9], Robert Trigwell[10]
Affiliations: Institute for Global Health, University College London, London, UK; Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK[1], Institute for Global Health, University College London, London, UK; Institute of Environment, Health and Societies, Brunel University, London, UK[2], Centre of Public Health Data Science, Institute of Health Informatics, University College London, London, UK[3], United Nations' Displacement Tracking Matrix, International Organization for Migration, International Organization for Migration, Juba, South Sudan[4], CU Population Center, Institute of Behavioral Science, University of Colorado Boulder Campus, Boulder, CO, USA; Faculty of Environmental Sciences, Technische Universität Dresden, Dresden, Germany[5], Health Management BD Foundation, Sector 6, Uttara, Dhaka, Bangladesh; Adjunct Faculty, Department of Public Health, North South University, Dhaka, Bangladesh[6], Department of Information Studies, University College London, London, UK[7], GMV Innovating Solutions Ltd, HQ Building, Thomson Avenue, Harwell Campus, Didcot, UK[8] WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK[9], United Nations' Displacement Tracking Matrix, International Organization for Migration, United Nations, London, UK[10]
Orcid ids: 0000-0002-5766-3986[1]
Contact e-mail: lydia.franklinos.16@ucl.ac.uk

Preprint statement: This article is a preprint and has not been peer-reviewed, under consideration and submitted to UCL Open: Environment Preprint for open peer review.
Funder: UCL Grand Challenges, UCL, London, UK
DOI: 10.14324/111.444/000042.v2
Preprint first posted online: 10 September 2021
Keywords: Big data, migration, cross-disciplinary research, policy, humanitarian, environment, displacement, climate change, health, data security, Environmental modelling, Public policymaking, Environmental justice and inequality/inequity, Statistics, Health and climate change

Institute of Global Health
University College London
30 Guilford Street
London
WC1N 1EH

18 June 2020

Dear Editorial Staff,

Key opportunities and challenges for the use of big data in migration research and policy

I am writing to submit the above manuscript for consideration for publication in UCL Open: Environment, as previously discussed with Ian Caswell and Professor Dan Osborn. Migration is high on the global political agenda, with one billion people recorded as having migrated in 2018 alone. Such levels of migration have great implications for global health, the global development agenda and political discourse at all levels of society. Meeting the needs of migrant populations requires accurate data on who is moving, where to and from, and what is driving their movements. However, current information on migrant populations is relatively scarce, leading to the UCL-Lancet Commission on Migration and Health to propose 'big data' (from mobile phone data to satellite data) as a potential solution to help address these knowledge gaps.

In response to the Commission's call, we held a cross-disciplinary workshop in London, UK in July 2019, bringing together UN representatives, humanitarian NGOs, policymakers and academics to facilitate knowledge exchange and identify the key opportunities and challenges for the implementation of big data in migration research. This workshop was supported by a UCL Grand Challenges grant for projects focusing on migration and displacement.

Here, we provide a summary of key discussion points and conclusions identified in the workshop to assist migration experts in deciding whether the use of big data is appropriate for their work. We also aim to stimulate discussion about the potential of this approach in aiding future migration research and policy and the needs of migrant populations globally, especially given ongoing high-level political narratives about migration, environmental drivers, and migrant health. Furthermore, the workshop was a highly interdisciplinary event and we take the opportunity to discuss how to develop further cross-disciplinary solutions to migration research.

I confirm that I have agreement from all authors to submit this paper and that the manuscript is not submitted elsewhere and is original.
I look forward to your response in due course.

Yours faithfully,

Dr Lydia Franklinos

Key opportunities and challenges for the use of big data in migration research and policy

Lydia H.V. Franklinos[1,2], Rebecca Parrish[1,3], Rachel Burns[4], Andrea Caflisch[5], Bishawjit Mallick[6,7], Taifur Rahman[8,9], Vasileios Routsis[10], Ana Sebastián López[11], Andrew J. Tatem[12], Robert Trigwell[13].

1. *Institute for Global Health, University College London, London, UK*
2. *Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK*
3. *Institute of Environment, Health and Societies, Brunel University, London, UK*
4. *Centre of Public Health Data Science, Institute of Health Informatics, University College London, London, UK*
5. *United Nations' Displacement Tracking Matrix, International Organization for Migration, International Organization for Migration, Juba, South Sudan*
6. *CU Population Center, Institute of Behavioral Science, University of Colorado Boulder Campus, Boulder, CO, USA*
7. *Faculty of Environmental Sciences, Technische Universität Dresden, Dresden, Germany*
8. *Health Management BD Foundation, Sector 6, Uttara, Dhaka, Bangladesh*
9. *Adjunct Faculty, Department of Public Health, North South University, Dhaka, Bangladesh*
10. *Department of Information Studies, University College London, London, UK*
11. *GMV Innovating Solutions Ltd, HQ Building, Thomson Avenue, Harwell Campus, Didcot, UK*
12. *WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK*
13. *United Nations' Displacement Tracking Matrix, International Organization for Migration, United Nations, London, UK*

Abstract

Migration is one of the defining issues of the 21st century. Better data is required to improve understanding about how and why people are moving, target interventions and support evidence-based migration policy. Big data, defined as large, complex data from diverse sources, is regularly proposed as a solution to help address current gaps in knowledge. The authors participated in a workshop held in London, UK, in July 2019, that brought together experts from the UN, humanitarian NGOs, policy and academia to develop a better understanding of how big data could be used for migration research and policy. We identified six key areas regarding the application of big data in migration research and policy: ethical and

security concerns around the use of big data; accessing and utilising data; integrating data sources and knowledge; understanding environmental drivers of migration; improving healthcare access for migrant populations; and addressing political narratives. We advocate the need for careful consideration of the challenges faced by the use of big data, as well as increased cross-disciplinary collaborations to advance the use of big data in migration research whilst safeguarding vulnerable migrant communities.

Keywords

Big data; migration; cross-disciplinary research; policy; humanitarian; environment; displacement; climate change; health; data security.

Introduction

With the number of global refugees reaching the highest levels since the second World War (1) and one billion migrants recorded in 2018 alone (2), human migration is high on the global political agenda. The UCL-Lancet Commission on Migration and Health (2) and the UN Global Compact on Migration (3) have called for improved data to understand drivers of migration, target interventions and support evidence-based migration policy. The application of big data in migration research and policymaking has been proposed as a possible solution to help address these knowledge gaps (2,4). Big data refers to large, complex data from varied sources, ranging from social media and mobile phone data (Figure 1), to electronic health records and satellite data, and has the potential to provide new sources of information for migration research (5,6). Previous studies have used big data to predict patterns of human movement during natural disasters (7) and track movement in near real-time (8), quantify migration at national scales (9–13), guide and evaluate humanitarian interventions (14) and examine the effects of human movement on disease transmission (15). In addition, satellite-based earth observation data has been used to map the relationship between environmental change and human movement (16,17), model subnational migration flows (18) and inform policy decisions (17,19,20). Despite the immense opportunity big data can provide for migration research and policy, several challenges have hindered its widespread implementation (2,4).

In response to the call for increased collaboration (2) and improved research on ways to utilise big data sources in the field of migration (4), we participated in a cross-disciplinary workshop in London, UK 3rd July 2019, bringing together UN representatives, humanitarian NGOs, policymakers and academics to facilitate knowledge-exchange and identify the key

opportunities and challenges for the implementation of big data in migration research. Here, we provide a summary of key discussion points identified in the workshop via presentations, panel discussions and break-out groups in which participants explored different topics and possible solutions. We provide major conclusions from the workshop supported by a review of relevant literature to assist migration experts in deciding whether the use of big data is appropriate for their work and to stimulate discussion about the potential of this approach in aiding migration research and policy, and the needs of migrant populations globally. In particular, the outcomes of this workshop may provide a timely resource for the recently launched Lancet Migration, a global collaboration of migration experts that aims to address evidence gaps and drive policy change in the field (21). Importantly, this workshop also tested a range of methods of collaborative working and identified challenges and opportunities in achieving a truly cross-disciplinary approach to migration research, an important but often neglected aspect in such a complex and politically charged topic as migration.

Opportunities and challenges relating to big data

The aims of the workshop were to facilitate discussion on both the opportunities of big data but also the challenges, shortfalls and current ethical considerations surrounding big data and its uses. Whilst it is beyond the scope of a single workshop or article to find solutions to these major challenges, the workshop proved a valuable forum for raising awareness and nuance discussion around these challenges which are often not fully recognised or addressed by researchers and practitioners alike when collecting or using big data as it relates to migration.

The format of the workshop consisted of keynote speeches, three panel discussions on the topics of 1) the role of big data in understanding the environmental drivers of migration, 2) the opportunities and challenges of big data in healthcare, and 3) the ethical considerations surrounding the use of big data in migration research and policy. The panel discussions were followed by group exercises which aimed to conceive structured, actionable solutions to a set of questions raised during the panel discussions. The resultant discussions and ideas were grouped into six topics which we summarise here and in Table 1.

Accessing and utilising big data

The first topic focused on the access, awareness and expertise required for big data use. The application of big data is often hindered by the fact that many big data sources such as mobile phones, internet-based platforms and other digital devices are managed by private companies who collect the data for business purposes. Therefore, costs associated with accessing big

data and issues of ownership are significant barriers to its use (22). Big data generation will vary geographically and may be reduced in many high mobility contexts where infrastructure (i.e., cell towers, wi-fi connection and electronic bank transfer services) is less established. Specific to the use of mobile phone call detail record (CDR) data, the presence of multiple mobile network operators in a country mean that multiple data sharing agreements would be required (23). In addition, there are significant issues around the potential extraction of sensitive information contained in big data (24,25) and data sources are often fragmentated across disciplines which reduces the awareness of available datasets (26). Accounting for multiple biases and the complex analyses required to interpret the data are further examples of methodological difficulties associated with the use of big data (6,27).

Workshop discussions highlighted the importance of understanding how, why and when the data were collected to identify potential gaps and biases, therefore ensuring it can be used effectively. There is great need for more centralised repositories of data, projects and publications such as The Humanitarian Data Exchange (28), to promote knowledge- sharing, collaborations and inform evidence-based programming. Increased partnerships between governments, international agencies, civil society, and the private sector are also required to improve data access and ensure the optimum exploitation of available data and technologies. Furthermore, capacity building in countries or organisations with an interest in big data analysis is needed to support cross-disciplinary research and improve specialist knowledge in certain regions. This could be achieved via collaborations with relevant partners and agencies such as demonstrated with the United Nations Economic Commission for Europe's (UNECE) Big Data Sandbox which provided a platform for statistical organisations to collaborate and learn to use big data analytics (29). However, there may be ethical considerations for private-public partnerships. For example, published commentaries have voiced fears over the partnership between the UN's World Food Programme and the data analytics company Palantir which may have serious consequences for the privacy and security of aid recipients due to the company's links to US intelligence agencies (30).

Integrating data sources and knowledge

The second topic concerned the integration of data and knowledge across disciplines. The main source of data for migration statistics originates from traditional methods such as household surveys recorded at local scales and national population estimates, as well as data on forced displacement collected through key informant networks (4) (Figure 2). Big data sources have the potential to complement traditional data and address significant spatial and temporal gaps via updating migration statistics in an accurate and low-cost way (4,10). For

example, analysis of CDR data can be used to replicate national internal migration statistics and complement outputs from censuses (10). However, integrating migration data from traditional methods with varied sources of big data requires new methodology that considers complex interactions over differing geographic and temporal scales. The quality of various data sets (e.g., demographic biases present in social media datasets) remains an unresolved challenge in teasing out comprehensive, policy-relevant results. Validating estimated migration using near-real-time big data is also problematic, with no trusted 'gold standard' currently available (31). The slow adoption of big data analyses in the humanitarian sector is partly due to a lack of expertise in how to apply these approaches in operational settings (32). Workshop participants discussed the need to bridge the gap between experts on the ground collecting the data via traditional methods and big data analysts via increased transdisciplinary training and collaborations. A recent workshop hosted by the International Organization for Migration (IOM) and the German Federal Foreign Office concluded that "greater cooperation and engagement among stakeholders" both within and external to the migration sector are required to inform decision making (33). If we are to integrate different data sources effectively, a collaborative cross-disciplinary approach is required to ensure we understand the data and how they can be used to deepen our understanding of the drivers and impacts of migration. This approach is practiced in "Data Collaboratives"; collaborative projects in which different sectors including private companies, research institutions, and government agencies collaborate to enable data exchange and help solve public problems (34). NetHope is an example of a Data Collaborative project which has helped to integrate data sources and produce maps of connectivity sites across Puerto Rico to assist in delivering aid in the aftermath of Hurricane Maria (35).

Understanding environmental drivers of migration

The third topic considered the use of big data in understanding and identifying environmental drivers of migration such as natural hazards and climate change (36) for example via remote sensing. Currently, there is no internationally agreed definition for "environmental migrant'' despite it being required to collect long-term data and guide the policies of governments and international agencies. The IOM have proposed a broad working definition (37) which importantly considers that environmental migration might be triggered both by sudden-onset events (38), such as earthquakes and cyclones, and slower environmental change processes, such as desertification and sea-level rise. In the context of slow-onset events and gradual environmental change, the effects on migration are often difficult to quantify since it can be hidden behind more immediate socioeconomic drivers such as poverty or political processes (39). There are opportunities for the application of big data to unpick these different drivers and to

help to define the term “environmental migrant’’. A further application could be in helping to quantify the effect of environmental change on the trapping of individuals or communities, usually due to rising poverty barriers which impede mobility. Human mobility can improve resilience and attenuate the negative outcomes of environmental degradation, but poverty, disability and social exclusion may limit people's ability to resort to migration as an adaptation strategy (40).

Despite the significant attention that environment-induced human migration has received (41), evidence on the relationship between migration and environmental change is limited. When used in combination with traditional datasets, big data has the potential to identify hot spots of environmental change and exposed populations that may be affected by the change and therefore liable to migrate, or to become trapped. Satellite data is a particularly valuable resource in analysis of environmental drivers as it enables the systematic, consistent and accurate monitoring of areas (even if remote or inaccessible) that are affected by anthropic or natural hazards. Indeed, satellite-based technologies are key to analysing observable effects of climate change effects and even predicting environmental-led migration (42-47). Furthermore, a study combined satellite data and CDR data to quantify the incidence, direction and duration of flood-driven migration, revealing important short-term (hours–weeks) aspects of migration associated with extreme weather events that are often not captured with traditional survey-based research (17). There is great potential for big data to help in understanding and addressing environment-related displacement and to inform policies that will improve resilience to environmental change and support migration that is required to improve the health and livelihoods of vulnerable people. In such examples, improved data on both environmental change (as a driver of migration) and of migration itself (such as displacement following a natural hazard) cannot ‘solve the problem’ of forced environmental migration but can inform interventions such as aid as well as discussions between affected communities and stakeholders in devising context-appropriate solutions for the future.

One of the most valuable aspects of satellite-based analyses is the capability for retrospective analysis which is required to detect changing patterns across space and time and to inform predictions. However, a recent review stated that current initiatives do not exploit the full possibilities of satellite-based earth observation in migration with a lack of services offering the systematic flow of detailed information to researchers, managers and migration analysts (48). One of the main gaps identified was that currently consolidated satellite-based monitoring systems work at regional scales which it is often too coarse to understand the specificities of how particular communities are affected, thus unable to inform sub-national policies design, implementation and monitoring. Indeed, reported discrepancies between the assumed

environmental drivers of migration perceived from satellite data analyses and the self-reported drivers of migration (e.g., see 49) underline the importance of including information on the lived experiences of migrants to inform actions.

Improving healthcare access for migrant populations

The fourth area of discussion focused on the potential for big data to enhance migrant health via improved disease outbreak preparedness, identification of vulnerable groups, increased access to healthcare and by informing evidence-based health interventions. Many studies have demonstrated the potential for big data to understand the impact of migration on the transmission of infectious diseases such as dengue (15), malaria (18,50) and cholera (51) at a national scale. These analyses highlight the potential benefit of big data use to improve preparedness and mitigation efforts for disease outbreaks. This may be particularly useful when supporting on-the-ground activities by helping to predict potential disease outbreaks for displaced persons. A recent collaboration between IOM and the mobile operator data analytics organisation Flowminder, is combining IOM Flow Monitoring Registry surveys with CDR data to gather anonymous information about people on the move at key transit points to inform public health interventions for the COVID-19 pandemic (52). A further application of big data in humanitarian settings is the use of satellite data to map refugee settlements (20) which can be used to ensure healthcare access for displaced persons.

A perennial issue in equitable healthcare access is identifying and addressing the needs of invisible communities, such as migrants, particularly those with an undocumented status (2). Shortfalls with traditional datasets and data collection processes such as semi-structured interviews and surveys means there is limited information on undocumented migrants and vulnerable groups (i.e., unaccompanied children, people with disabilities and members of the LGBTI community). There is great potential for big data address the paucity of information on these groups, their health, access to healthcare and differing healthcare needs. In addition, it was suggested that the healthcare needs of migrants settled in countries such as the UK could also be improved by big data analysis, for example via a general migrant longitudinal study such as the cohort studies performed by the UK Economic and Social Research Council (53). There are also vast applications for big data in implementing evidence-based health interventions that need to be explored, specifically in identifying differences in patient responses to treatments and tailoring healthcare to the specific needs of individuals (54).

The COVID-19 pandemic has received widespread support for the use of big data in disease surveillance systems globally (55). Such approaches have had various levels of success at

curbing infection rates, particularly crucial for vulnerable persons (which could include some migrant communities). However, if the use of such big data approaches in healthcare became common place, it may cost many the protection that invisibility currently offers, with many communities such as undocumented migrants fearing disproportionate effects of state surveillance (56). Such ethical conundrums remain unresolved and often underexplored. At best, it is likely that big data should not be seen as a perfect solution but as one optional tool within a wider social toolkit which retains traditional and non-digital interventions.

Ethical, privacy and security concerns

The fifth topic focused on ethical, privacy and security concerns regarding the use of big data in migration research. This topic proved to be cross-cutting, with themes re-emerging across other topics. Principally, a recurrent question arose about whether it is appropriate to collect and/or use sensitive data in the pursuit of greater understanding for researchers and policymakers. The collection of personal data including migrant status is a contentious issue. There are concerns that information on personal migration status may create or increase existing discriminatory practices in society such as provision of healthcare and access to state funds, or that mobile tracking devices may be used against a migrant to forcibly return them to a previous location (2). Additionally, various sources of big data (such as social media) are consumed differently based on geography,demography and access. As such, any analysis of these datasets will carry these biases, which may follow through into policies. This may result in policies which propagate stereotypes and discriminatory practices, or else continue to underserve invisible groups (e.g., those not engaged with social media or with smaller social networks) (31).

Ethics in the context of big data in migration may be considered in several ways. First, it may relate to the way in which the research is conducted and whether there has been consideration for data privacy and security. Secondly, it could refer to decision-making regarding migrants with consideration for their lived experience, especially in humanitarian situations. A recent report on migration noted that discussions on ethics often focus on the legal aspects of data protection rather than understanding how the results of analyses may detrimentally impact affected populations and counter the humanitarian principles to “do no harm” (33). Furthermore, data protection measures are often focused on personal data (e.g., General Data Protection Regulation in the EU) and do not consider group data protection needed to work with vulnerable groups (57).

It is important to consider who benefits from the use of big data sources in migration research. At the individual level, migrants may not wish for additional data to be gathered about them and may perceive no benefits of the process (58). However, at the community level, such data and analysis may help to address the health needs of migrants more generally. Certainly, there will be benefits to the academic community seeking to study migrant health needs and to decision-makers seeking evidence-based solutions. Pursuit of these research and policy goals can result in overlooking the individual rights and raise ethical issues for many vulnerable people (58). Furthermore, forcibly displaced people fleeing persecution may have little trust in authorities and therefore be less willing to seek healthcare or consent to having their data collected. This creates a barrier for healthcare professionals, humanitarian workers and researchers who wish to respect the rights of the individual, whilst deriving a better understanding of migration pathways and healthcare needs. Workshop discussions highlighted the power imbalance between various parties involved; those seeking data including governments and academics often from the Global North with inherent biases and power, and those the data is being sought from who are often vulnerable persons in precarious or dangerous situations, many originating from the Global South (59). Evidence of this power imbalance is seen with 'high-risk experiments' with new technologies such as the use of Canada's automated decision-making technology in immigration and refugee applications (60) which often lack regulation. Even after applying advanced safeguarding practices and aggregated outputs, researchers may still be reluctant to apply big data analysis for migration research as policy- makers often have their own agendas and may use the methods and deliverables in ways not intended or anticipated by the research authors.

Another established concern regarding the ethical deployment of big data in migration research is the focus on pattern-based analyses, rather than on achieving a conceptual interpretation and critical analysis for why behavioural trends might emerge and under what set of circumstances and assumptions (61,62). Indeed, the limitations and bias of big data remains an under-explored aspect in the discourse around ethical implications of big data. The COVID-19 pandemic provides a pertinent example of how the lack of data from vulnerable communities such as refugees or people on the move has led to the underrepresentation of these communities in the narrative and political responses of the pandemic (63).

Addressing political narratives

The final topic concerned the role of big data in high level political narratives around migration. With ongoing antimigrant rhetoric existing at all levels of government and society, migrant

research has the opportunity and mandate to address such political narratives. There are many examples of authorities treating migrants as political pawns or as statistical figures (64). Therefore, it is imperative that big data is not used to further discriminate against migrant communities or to target certain populations, but rather to support inclusive and fair migration governance. This can be particularly problematic with big datasets, where assumptions have been made throughout the data collection and analysis process. For example, in 2017 the UK data analytics company CGI together with the Dutch statistical agency CBS, conducted a study into migration forecasting of people in Syria using Twitter data (62). The study required a set series of assumptions to be made about Twitter content (more specifically its English translation) and was unable to consider the context within which tweets were posted. This case study demonstrates how the promotion of big data can reify such assumptions and distort the original meaning or intent behind an individual's migration decisioning, as well as pervert the aggregate narrative behind migration patterns (62). Conversely, an positive example discussed within the workshop was the Sentinel project; an NGO that works to gather and disseminate ' trusted' information to local people and governments in order to counter the spread of misinformation, antimigrant rhetoric and to prevent resultant hate crime and genocides (65). Participants also deliberated whether the increased evidence provided by big data would be instructive and powerful enough to overcome political and social biases associated with the topic. Given the highly political nature and high stakes of migration policy for migrants as well as for governments and the international community, more evidence may alone be insufficient to achieve multilateral, progressive action. It is worth therefore considering other factors contributing to the political discourse and designing strong, cross-disciplinary communication tactics to support maximum impact of evidence. Furthermore, it is worth considering how paucity in migration data has helped to shape public and media perceptions of global migration patterns to date, and whether big data could be used to address these perceptions.

Discussion

The application of big data in migrant research shows much promise in addressing the current gaps in knowledge. Big data sources can help to update internal migration statistics by addressing the significant gaps in quantity and quality of data collected from traditional methods (4,10). When combined with field-level data derived from household surveys and key-informant networks, big data can be used to detect how sudden onset natural hazards and gradual environmental change (e.g., desertification and sea-level rise associated with climate change) impact migration patterns. This can help to inform planning and scenario

building, as well as contributing to a more comprehensive definition of “environmental migrant” which is critical for migration policy within the context of ongoing environmental change. In addition, it has a potential application in considering the differing healthcare needs of migrants as well as identifying vulnerable populations unable to migrate due to environmental change. There is also scope for big data to inform evidence-based health interventions for migrant populations in everyday and emergency (including displacement) settings. Yet despite the vast opportunities that big data present, there are some important areas to consider before using these varied and complex data sources. Increased cross-disciplinary partnerships are required to improve data access, knowledge-sharing and capacity building across sectors and regions. In addition, a collaborative cross-disciplinary approach is required to ensure the different datasets are understood and to develop new methodologies to integrate data sources and identify complex interactions that influence how and why people are moving. Furthermore, the reported lack of agreement within the humanitarian sector on how migration modelling should be applied needs to be addressed so analyses can be effective (33). It is important to challenge the assumption that big data is always a suitable and insightful tool to use in research for migration policy. Whilst tempting to consider big data the magic bullet, it may not always be appropriate or may need to be used in conjunction with traditional methods.

International legislation is required to sufficiently address how migrant data should be collected and used to ensure ethical conduct by data gatherers and owners and the safeguarding of human rights, even in sensitive migration contexts (66). The United Nations Development Group (66) and Office for the Coordination of Humanitarian Affairs (67) provide guidelines for safe and ethical data management in humanitarian situations however, there is no legal enforcement of these practices. Although researchers would like quicker and easier access to data, workshop discussions challenged whether the process should be hastened, suggesting that administrative obstacles force researchers to duly consider whether additional data is necessary and beneficial to the current state of knowledge, given the risks and trade-offs that must be made. A key output of the workshop was a consensus that researchers and decision-makers must first ask why they require additional data and whether this is what all parties, particularly migrants, would want. As well as data protection issues, it is also imperative to understand the potential harmful impact of analyses on vulnerable migrant populations (33). It is especially important that big data is not used to further discriminate or target migrant populations considering current antimigrant political narratives.

In pursuit of cross-disciplinary collaboration, the workshop brought together a range of representatives from the UN, government, humanitarian agents and academics from a range of backgrounds. Cross-sector engagement in the workshop was difficult to achieve which may

be due to differences in the objectives of different sectors, as well as the language of engagement used. For instance, humanitarian organisations were particularly difficult to engage, and it is thought that this was due to both a shortage of networks linking academia and humanitarian organisations as well as differences in short, medium and long-term needs and objectives of the two sectors. We trialed different methods to stimulate interdisciplinary work including the use of business canvases (68) to explore and present solutions to questions. This approach was useful for stimulating debate within the groups and producing well-considered outputs. However, future interdisciplinary events would benefit from the development of methods that consider the language styles and information sharing techniques of different disciplines and thus facilitate effective communication and knowledge-sharing (4,33,69,70). The workshop also elucidated the extent to which the direction of conversation is steered by which actors are engaging in the conversation. For example, actors engaging with migrants directly (i.e., service providers) focus on practical implications. This serves as a pertinent reminder of the importance for academics to create and utilise diverse networks as well as the need for actors with gatekeeping power to exercise due diligence and engage a wide range of stakeholders and interest groups in discussions. Overall, the workshop highlighted the benefits of cross-disciplinary work, enabling the identification of key topics from a variety of angles and providing meaningful and effective outputs. Furthermore, we hope this workshop assists in cultivating a future transdisciplinary approach to migration research, whereby there is a move beyond the collaboration of individual disciplinary perspectives to develop curriculum integration that organizes knowledge production in the context of real-world problems (71).

462 **Table 1.** Key questions, challenges, opportunities and solutions for the use of big data in migration research and policy.
463

Topics	Key research and policy questions	Potential challenges	Potential opportunities and solutions
Accessing and utilising big data	● How can we improve access to big data sources? ● How can we enhance awareness of available data? ● How can we develop the expertise required to use big data across disciplines?	● Issues of ownership and costs in accessing big data are significant barriers. ● If multiple mobile network operators are operating in a country, multiple data sharing agreements would be required (23). ●Geographic biases in the type of data available. ● Poor awareness of fragmented data sources. ● Issues around the potential extraction of sensitive information contained in big data, therefore access arguably should be difficult for these data.	● Opportunities for the development of centralised repositories of data such as The Humanitarian Data Exchange (28) to promote collaboration and knowledge-sharing. ● Opportunities for partnerships across different sectors to improve access to available data and technologies. ● Capacity building projects across disciplines such as the UNECE Big Data Sandbox (29) are needed to support big data use for migration research.
Integrating data sources and knowledge	● How can we produce more detailed and recent migration statistics using big data? ● How can we best integrate data from different sources? ● How do we manage fragmented	● Complex analyses are required to account for multiple biases in different datasets. ● Lack of awareness and big data expertise in humanitarian sector has led to its slow adoption.	● The use of big data sources combined with traditional survey-based research can reveal important aspects of migration that are often not captured (e.g., CDRs reveal short-term migration patterns) (52). ● Potential for methodological innovation for the

	data sources across varied spatial and temporal scales? ● How can we develop a collaborative cross-disciplinary approach to address the challenges in the field of migration?		integration of different data and the development of a ‘gold standard’ to estimate migration using near-real-time big data. ● Opportunities for cross-disciplinary “Data Collaboratives" (34) to enable data exchange and help address complex problems.
Understanding environmental drivers of migration	● How can big data be used to assess the ongoing impact of climate change on migration? ● Can big data help to identify populations that are vulnerable to environmental change? ● How can big data be used to predict mass migration events due to environmental change?	● Satellite-based environmental data is often at coarse geographic scales which are not suitable to inform sub-national policies and actions. ● There may be discrepancies between the assumed environmental drivers of migration perceived from satellite data analyses and the self-reported drivers of migration.	● Opportunity to address the lack of evidence on the relationship between migration and environmental change and to help to define the term “environmental migrant’ that is needed for policy action. ● Potential to reveal important aspects of migration associated with extreme weather events that are often not captured with traditional data. ● Potential to unpick socioeconomic factors that may be limiting people's ability to migrate as an adaptation strategy to environmental change. ● Opportunities to inform policies that will improve resilience and support migration associated with environmental change.
Improving healthcare access for migrant populations	● How can big data be used to address the immediate health needs of displaced persons in camps?	● Big data approaches in healthcare risk jeopardise the protection from state surveillance that is currently provided to vulnerable communities such as	● Big data can help to understand the impact of migration on the transmission of infectious diseases. ● Potential to support on-the-ground activities,

	● How can big data help us learn more about undocumented migrants, their health and healthcare needs? ● How can big data be used to implement evidence-based health interventions?	undocumented migrants.	helping to address the immediate health needs of displaced persons and to predict disease outbreaks. ● Potential for big data to improve equitable healthcare access for underrepresented communities, such as undocumented migrants and vulnerable groups. ● Opportunities to improve the specific healthcare needs of migrants with settled status. ● Applications for big data in identifying differences in patient responses to treatments and tailoring healthcare to the specific needs of individuals.
Ethical, privacy and security concerns	● What is meant by ethics in the context of big data in migration research? ● Who benefits from the use of big data (migrants at individual or community level, academic community, policymakers)? ● How do power imbalances influence the use of big data? ● How can we achieve ethical data usage?	● Big data analytics may lead to the introduction of stakeholders which have varied motivations. ● Biases in big data may follow through into policies, propagating stereotypes and discriminatory practices, or else continuing to underserve invisible groups. ● There is often focus on the legal aspects of data protection for personal data rather than the potential negative impacts on affected vulnerable groups.	● Ethical considerations and safeguarding practices are required when involving varied stakeholders in big data analytics ● The use of big data in migrant research requires the clear stating of assumptions and methods to correct for biases to prevent the propagation of biases in policy. ● Ethical considerations and safeguarding practices are required when involving varied stakeholders in big data analytics. ● Researchers and decision-makers must

		● Big data use in migration can promote the power imbalance between those seeking data and those the data is being extracted from. ● There is no legal enforcement of guidelines for safe and ethical data management in humanitarian situations.	critically examine and justify why they require the use of big data and whether this is what all parties, particularly migrants, would want.
Addressing political narratives	● How can we prevent the use of big data for the discrimination of certain populations? ● What role could big data have in addressing the negative political narratives around migration?	● Big data use may lead to lack on nuance behind the motivations for migration, potentially distorting the narrative behind migration patterns. ● Ongoing anti-immigration rhetoric means it is imperative that big data is not used to further discriminate against migrant communities or to target certain populations.	● Opportunity to help to address negative political narratives and to support inclusive and fair migration governance as seen with the Sentinel project which works to counter the spread of misinformation and antimigrant rhetoric (65). ● Designing strong, cross-disciplinary communication tactics to support maximum impact of evidence. ● Opportunity to help to challenge public and media perceptions of global migration which has been shaped by misinformation due to paucity in migration data.

Figure 1. Key statistical indicators for global big data use. The number of users and portion of the population that has access (penetration) to the internet, mobile phones, social media and mobile social media. Data were accessed via the Global Digital Report 2019 (72).

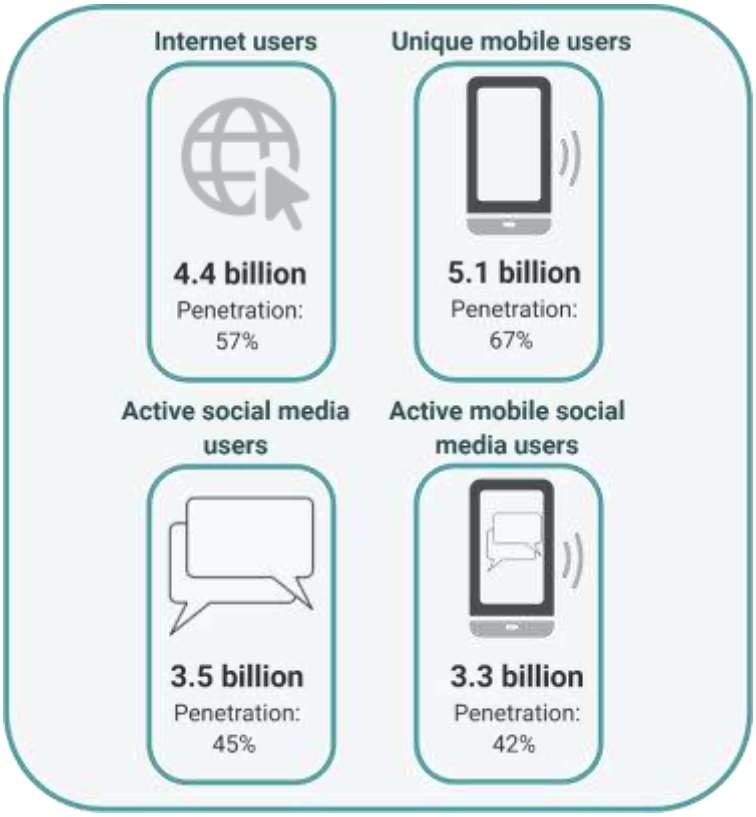
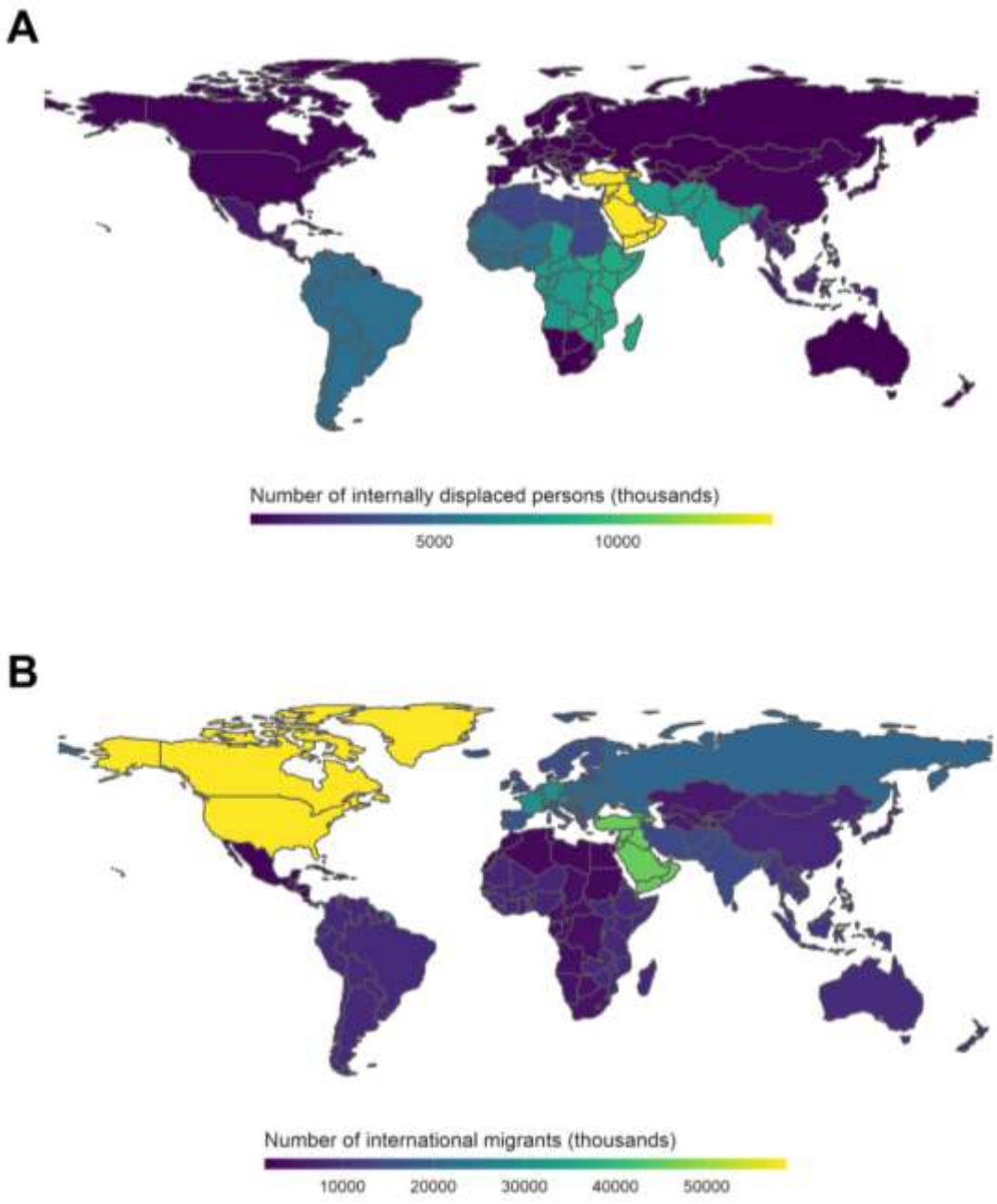


Figure 2. Global human migration by subregion. A) Number of internally displaced persons (from conflicts and disasters) in 2020 ranging from 22 in the Eastern Europe (dark blue) to 14,200,000 in Western Asia (yellow); data were accessed via the Internal Displacement Monitoring Centre' s Global Internal Displacement database (73). B) Number of international migrants in 2019 ranging from 1,542,000 in the Caribbean (dark blue) to 58,647,000 in North America (yellow); data were accessed via the UN Department of Economic and Social Affairs, Population Division database (74). The legends for both graphs show the number of migrants in thousands.



480

481

Funding and acknowledgements

The workshop was supported by a UCL Grand Challenges grant (506002-100-156425). We thank Leroy Clement, Ilan Kelman, Bayes Ahmed, Chris Langridge, Yasna Palmeiro-Silva, Laura Busert, Maheswar Satpathy, Emeline Rougeaux and Alicia Matthews for their help in the organisation of the workshop.

Author contributions

LF and RP led the workshop and took the lead role in writing the report. All authors participated in discussion sessions at the workshop and contributed to the report. Authors 3–10 are listed alphabetically.

Declarations and conflicts of interest

All other authors declare no conflicts of interest in connection to this article.

Open data and data availability

No further data were used in addition to referenced works.

References

1. UNHCR. Global Trends: Forced displacements in 2015. Geneva, Sw; 2015.

2. Abubakar I, Aldridge RW, Devakumar D, Orcutt M, Burns R, Barreto ML, et al. The UCL–Lancet Commission on Migration and Health: the health of a world on the move. The Lancet. 2018 Dec 15;392(10164):2606–54.

3. UN. Global Compact for Safe, Orderly and Regular Migration. 2018.

4. IOM. Data bulletin-big data and migration. Ispra: International Organization for Migration; 2018. Available from: https://publications.iom.int/system/files/pdf/issue_5_big_data_and_migration.pdf [Accessed 2020 Mar 20]

5. Fleming L, Depledge M, Leonelli S, Gordon-Brown H, Leonardi G, Golding B, et al. Big data in environment and human health. Oxf Res Encycl Environ Sci. 2017;July:1–27.

6. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med. 2013;10(4):2–5.

7. Lu X, Bengtsson L, Holme P. Predictability of population displacement after the 2010 Haiti earthquake. Proc Natl Acad Sci. 2012; 109(29):11576-11581

8. Wilson R, Erbach-Schoenberg E zu, Albert M, Power D, Tudge S, Gonzalez M, et al. Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal earthquake. PLOS Curr Disasters. 2016 Feb 24

9. Lu X, Wrathall DJ, Sundsøy PR. Detecting climate adaptation with mobile network data in Bangladesh: anomalies in communication, mobility and consumption patterns during cyclone Mahasen. Clim Change. 2016;1–15.

10. Lai S, Erbach-Schoenberg E zu, Pezzulo C, Ruktanonchai NW, Sorichetta A, Steele J, et al. Exploring the use of mobile phone data for national migration statistics. Palgrave Commun. 2019 Mar 26;5(1):1–10.

11. Ruktanonchai NW, Ruktanonchai CW, Floyd JR, Tatem AJ. Using Google Location History data to quantify fine-scale human mobility. Int J Health Geogr. 2018 Jul 27;17

12. Spyratos S, Vespe M, Natale F, Weber I, Zagheni E, Rango M. Quantifying international human mobility patterns using Facebook Network data. PLOS ONE. 2019 Oct 24;14(10):e0224134.

13. Palotti J, Adler N, Morales-Guzman A, Villaveces J, Sekara V, Herranz MG, et al. Monitoring of the Venezuelan exodus through Facebook's advertising platform. PLOS ONE. 2020 Feb 21;15(2):e0229175.

14. Peak CM, Wesolowski A, zu Erbach-Schoenberg E, Tatem AJ, Wetter E, Lu

X, et al. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. Int J Epidemiol. 2018 Oct 1;47(5):1562–70.

15. Wesolowski A, Qureshi T, Boni MF, Roe P, Johansson MA, Basit S. Impact of human mobility on the emergence of dengue epidemics in Pakistan. PNAS. 2015;112(38):11887–11892.

16. Van der Geest K, Vrieling A, Dietz T. Migration and environment in Ghana: a cross-district analysis of human mobility and vegetation dynamics. Environ Urban. 2010;22(1):107–123.

17. Lu X, Wrathall DJ, Roe P, Wetter E, Iqbal A, Qureshi T, et al. Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. Glob Environ Change. 2016;38:1–7.

18. Sorichetta A, Bird TJ, Ruktanonchai NW, zu Erbach-Schoenberg E, Pezzulo C, Tejedor N, et al. Mapping internal connectivity through human migration in malaria endemic countries. Sci Data. 2016 Aug 16;3(1):1–16.

19. Hauer ME, Fussell E, Mueller V, Burkett M, Call M, Abel K, et al. Sea-level rise and human migration. Nat Rev Earth Environ. 2020 Jan;1(1):28–39.

20. Logar T, Bullock J, Nemni E, Bromley L, Quinn JA, Luengo-Oroz M. PulseSatellite: A tool using human-AI feedback loops for satellite image analysis in humanitarian contexts. ArXiv200110685 Cs Eess. 2020 Jan 28

21. Orcutt M, Spiegel P, Kumar B, Abubakar I, Clark J, Horton R. Lancet Migration: global collaboration to advance migration health. The Lancet. 2020 Feb 1;395(10221):317–9.

22. Kirkpatrick R, Vacarelu F. A decade of leveraging big data for sustainable development. UN Chronicl. 2018;3(December). Available from: https://unchronicle.un.org/article/decade-leveraging-big-data-sustainable-development [Accessed 2020 Feb 11]

23. Gerstle T, Hardimon M, Hunt G, Main I, Paudel L, Pontón D, et al. Assessing the use of Call Detail Records (CDR) for monitoring mobility and displacement. Feb 2021. Princeton, NJ, USA: Princeton University School of Public and International Affairs. Available from: https://www.migrationdataportal.org/sites/default/files/2021-02/IOM_Princeton_CDRReport_Feb2021_web.pdf [Accessed 2021 Sep 9]

24. von Mörner M. Application of call detail records - chances and obstacles. Transp Res Procedia. 2017;25:2233–41.

25. de Montjoye Y-A, Gambs S, Blondel V, Canright G, de Cordes N, Deletaille S, et al. On the privacy-conscientious use of mobile phone data. Sci Data.

2018 Dec 11;5(1):1–6.

26. Verhulst SG, Young A. The potential and practice of data collaboratives for migration. Stanford Social Innovation Review. 2018. Available from: https://ssir.org/articles/entry/the_potential_and_practice_of_data_collaboratives_for_migration [Accessed 2020 Mar 19]

27. Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. The impact of biases in mobile phone ownership on estimates of human mobility. J R Soc Interface. 2013 Apr 6;10(81):20120986

28. Welcome - Humanitarian Data Exchange. Available from: https://data.humdata.org/ [Accessed 2020 Jun 18]

29. Vale S. Big Data Sandbox. UNECE. 2016. Available from: http://www1.unece.org/stat/platform/display/bigdata/Sandbox [Accessed 2020 Mar 19]

30. Parker B. New UN deal with data mining firm Palantir raises protection concerns. The New Humanitarian. 2019. Available from: https://www.thenewhumanitarian.org/news/2019/02/05/un-palantir-deal-data-mining-protection-concerns-wfp [Accessed 2020 Mar 19]

31. Sîrbu A, Andrienko G, Andrienko N, Boldrini C, Conti M, Giannotti F, et al. Human migration: the big data perspective. Int J Data Sci Anal. 2021;11:341–360 (2021). https://doi.org/10.1007/s41060-020-00213-5

32. Oroz ML. From big data to humanitarian-in-the-loop algorithms. UNHCR Innovation. 2018. Available from: https://www.unhcr.org/innovation/big-data-humanitarian-loop-algorithms/ [Accessed 2020 Feb 11]

33. IOM. Workshop report on forecasting human mobility in contexts of crises. Berlin: German Federal Foreign Office (FFO) and the International Organization for Migration (IOM); 2019 Oct

34. Winowatan M. The emergence of data collaboratives...in numbers. The Governance Lab. 2018. Available from: http://thegovlab.org/the-emergence-of-data-collaboratives-in-numbers/ [Accessed 2020 Mar 20]

35. NetHope Blog. Unlocking insights from data: Collaboration with private sector creates cutting-edge maps for disaster response – NetHope. 2018. Available from: https://nethope.org/2018/09/10/unlocking-insights-from-data-collaboration-with-private-sector-creates-cutting-edge-maps-for-disaster-response/ [Accessed 2020 Mar 20]

36. Martin S, Ferris E, Kumari K, Bergmann J. The global compacts and environmental drivers of migration. KNOMAD; 2018. Report No.: Policy Brief 11.

37. IOM. International Migration Law: Glossary on migration. Geneva: International Organization for Migration; 2019. Available from: https://www.iom.int/glossary-migration-2019 [Accessed 2020 Mar 30]

38. Disaster Displacement. Key Definitions. Platform on Disaster Displacement. 2020. Available from: https://disasterdisplacement.org/the-platform/key-definitions [Accessed 2020 Mar 30]

39. Ionesco D, Mokhnacheva D, Gemenne F. The Atlas of Environmental Migration. 1st ed. Oxon, UK: Routledge; 2017.

40. Oakes R, Banerjee S, Warner K. Chapter 9: Human mobility and adaptation to environmental change. In: World Migration Report 2020. Geneva: United Nations University Institute for Environment and Human Security, International Organization for Migration and UNFCCC Secretariat.; 2019

41. Maystadt J, Mueller V. Environmental migrants: A myth? International Food. Policy Research Institute, Research Brief 18; 2012. Available from: https://www.ifpri.org/publication/environmental-migrants [Accessed 2021 Aug 20]

42. Chen J, Mueller V. Coastal climate change, soil salinity and human migration in Bangladesh. Nature Clim Change. 2018;8, 981–985.

43. Chen J, Mueller V, Jia Y, Kuo-Hsin Tseng S. Validating migration responses to flooding using satellite and vital registration data. Amer Econ Rev 2017;107 (5): 441–45.

44. Liu, Zhen, Balk, D. Urbanisation and differential vulnerability to coastal flooding among migrants and nonmigrants in Bangladesh. Popul Space Place. 2020; 26:e2334.

45. Mueller V, Gray C, Kosec K. Heat stress increases long-term human migration in rural Pakistan. Nat Clim Change. 2014 Mar;4(3):182–5.

46. Davis KF, Bhattachan A, D'Odorico P, Suweis S. A universal model for predicting human migration under climate change: examining future sea level rise in Bangladesh. Environ Res Lett. 2018 Jun;13(6):064030.

47. DTM. South Sudan: seasonal floods analysis. International Organization for Migration; 2019 Oct. Available from: /reports/south-sudan-%E2%80%94-seasonal-flooding-maps-november-2019 [Accessed 2020 Mar 30]

48. Lang S, Füreder P, Riedler B, Wendt L, Braun A, Tiede D, et al. Earth observation tools and services to increase the effectiveness of humanitarian assistance. Eur J Remote Sens. 2019 Oct 30;0(0):1–19.

49. Hoffmann EM, Konerding V, Nautiyal S, Buerkert A. Is the push-pull paradigm useful to explain rural-urban migration? A case study in Uttarakhand, India. PLoS One. 2019;14(4):e0214511.

50. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee

CO. Quantifying the impact of human mobility on malaria. Science. 2012 Oct 12;338(6104):267-70.

51. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, Rebaudet S, Piarroux R. Using mobile phone data to predict the spatial spread of cholera. Sci Rep. 2015 Mar 9;5: 8923.

52. Harrison, J, M Harper, J Gray, and V Lefebvre. Privacy-Conscious Data Analytics to Support the COVID-19 Response in Haiti. Report 2. Jun 2020. Stockholm, Sweden: Flowminder Foundation. Available from: https://haiti.iom.int/sites/haiti/files/documents_files/Haiti_Digicel_DTM_Slide_Report_June2020_eng.pdf [Accessed 2021 Sep 9]

53. ESRC. Centre for Longitudinal Studies. UKRI - Economic and Social Research Council. 2020. Available from: https://esrc.ukri.org/research/our-research/centre-for-longitudinal-studies/ [Accessed 2020 Mar 17]

54. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health. 2019 Oct 1;29(Supplement_3):23-27.

55. Zwitter A, Gstrein OJ. Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. Int J Humanitarian Action. 2020;5(1):4.

56. Pelizza A, Lausberg Y, Milan S. The dilemma of making migrants visible to COVID-19 counting. Processing Citizenship Blog. April 28, 2020. https://processingcitizenship.eu/the-dilemma-of-making-migrants-visible-to-covid-19-counting/ [Accessed 2020 Aug 20]

57. World Economic Forum. Civil Society in the Fourth Industrial Revolution: Preparation and Response. Cologny/Geneva; 2019

58. Lamber R, Pinter K, Aigner A, Reiterer M, Kappel K, Grechenig T. Ethical issues arising through identification and registration systems applied in a European refugee camp. In: 2019 9th International Conference on Advanced Computer Information Technologies (ACIT). 2019. p. 320–4.

59. Nawyn SJ. Migration in the Global South: exploring new theoretical territory. Int J Sociol. 2016 Apr 2;46(2):81–4.

60. Molnar, 2020. AI and Migration Management. In: The Oxford handbook of ethics of AI, Eds. Dubber, Pasquale, Das, Oxford handbooks online. 2020. doi:10.1093/oxfordhb/9780190067397.013.49.

61. Halford S, Savage M. Speaking sociologically with big data: symphonic social science and the future for big data research. Sociology. 2017;51(6):1132-1148.

62. Taylor L, Meissner F. A crisis of opportunity: market-making, big data, and the consolidation of migration as risk. Antipode. 2020;52(1):270-290.

63. Milan S, Treré E. The rise of the data poor: the COVID-19 pandemic seen

from the margins. Soc Media Soc. 2020 Aug 11;6(3):2056305120948233.

64. Arcimaviciene L, Baglama SH. Migration, metaphor and myth in media representations: the ideological dichotomy of “them” and “us”: SAGE Open. 2018 May 9

65. The Sentinel Project. The Sentinel Project. 2018. Available from: https://thesentinelproject.org/ [Accessed 2020 Mar 20]

66. UNDG. Data Privacy, Ethics and protection: guidance note on big data for achievement of the 2030 Agenda. United Nations Development Group; 2017

67. The Centre for Humanitarian Data. Data Responsibility Guidelines. 2019.

68. Osterwalder A, Pigneur Y. Business model generation. Amsterdam: Self Published; 2009.

69. Gooch D, Vasalou A, Benton L. Impact in interdisciplinary and cross-sector research: Opportunities and challenges. J Assoc Inf Sci Technol. 2017;68(2):378–91

70. Vogel KM, Tyler BB. Interdisciplinary, cross-sector collaboration in the US Intelligence Community: lessons learned from past and present efforts. Intell Natl Secur. 2019 Sep 19;34(6):851–80.

71. Choi BCK, Pak AWP. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. Clin Investig Med Med Clin Exp. 2006 Dec;29(6):351–64.

72. We Are Social, Hootsuite. Global Digital Report 2019. New York; 2019. Available from: https://wearesocial.com/global-digital-report-2019 [Accessed 2020 Apr 1]

73. Internal Displacement Monitoring Centre (IDMC). Global Internal Displacement Database Data. 2020. https://www.internal-displacement.org/database/displacement-data [Accessed 2020 Aug 25]

74. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019, Online Edition. Rev. 1. 2019.