



ETHICAL CONSIDERATIONS IN RE-USING PRIVATE SECTOR DATA FOR MIGRATION-RELATED POLICY

A practitioners' perspective



IOM
UN MIGRATION

GLOBAL DATA
INSTITUTE

The opinions expressed in the report are those of the authors and do not necessarily reflect the views of the International Organization for Migration (IOM). The designations employed and the presentation of material throughout the report do not imply the expression of any opinion whatsoever on the part of IOM concerning the legal status of any country, territory, city or area, or of its authorities, or concerning its frontiers or boundaries. IOM is committed to the principle that humane and orderly migration benefits migrants and society. As an intergovernmental organization, IOM acts with its partners in the international community to: assist in meeting the operational challenges of migration; advance understanding of migration issues; encourage social and economic development through migration; and uphold the human dignity and well-being of migrants.

This publication has been issued without formal editing by the Publications unit of IOM.

CONTENT

Sievers, Niklas; Griesmer, Liz; Rango, Marzia; Trigwell, Robert; Jusselme, Damien.

LAYOUT DESIGN AND DATA VISUALIZATION

Roberta Aita

Global Migration Data Analysis Centre (GMDAC) International Organization for Migration (IOM)

Taubenstr. 20–22 D 10117 Berlin, Germany

Tel.: +49 30 278 778 22 Fax: +49 30 278 778 99

Please visit the GMDAC website for publications, resources, and events: <http://gmdac.iom.int>.

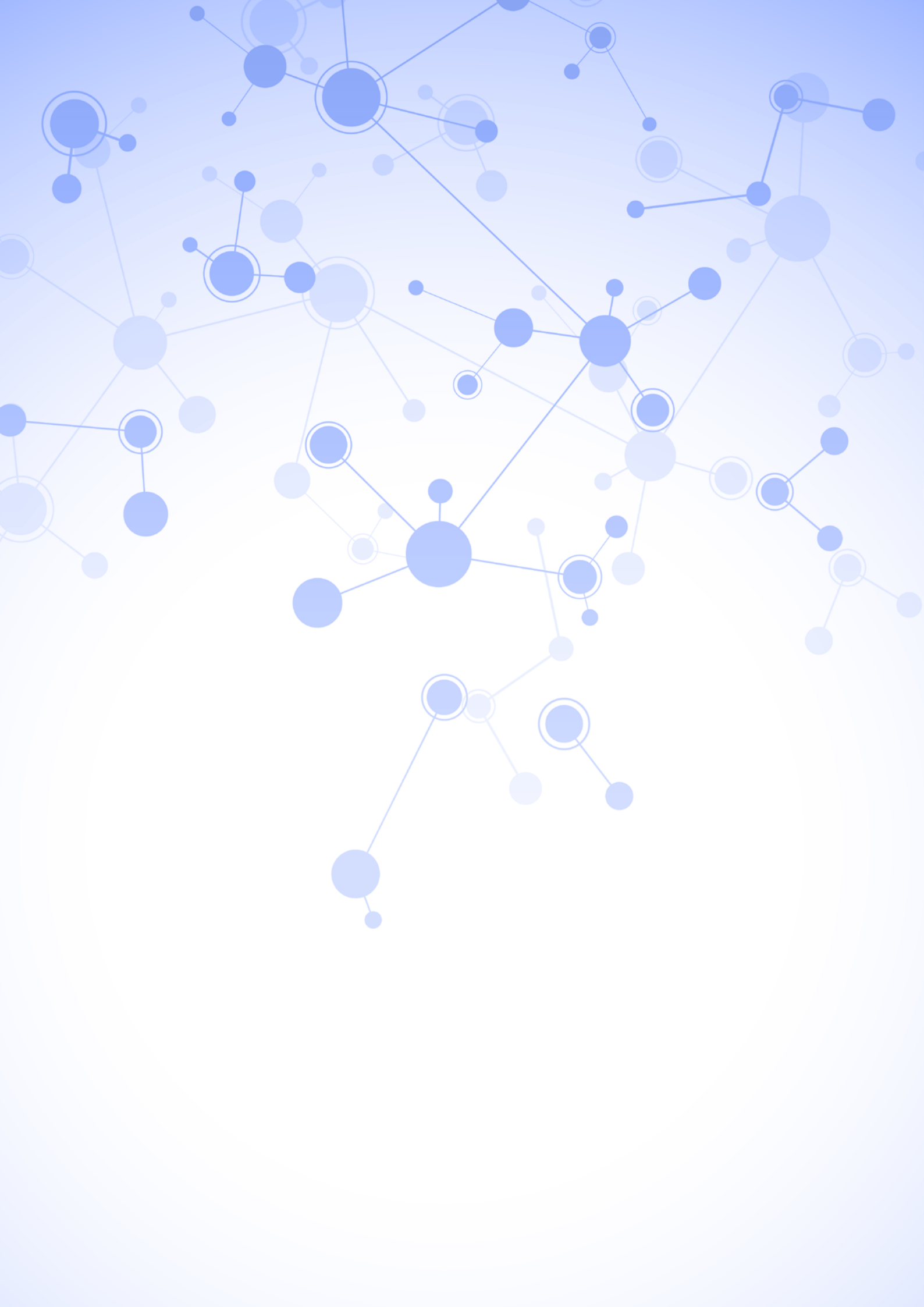
Suggested citation:

Sievers, Niklas; Griesmer, Liz; Rango, Marzia; Trigwell, Robert; Jusselme, Damien. 2022. *Ethical Considerations in Re-Using Private Sector Data for Migration-related Policy: A Practitioners' Perspective*.

ETHICAL CONSIDERATIONS
IN RE-USING PRIVATE SECTOR DATA
FOR MIGRATION-RELATED POLICY

A practitioners' perspective





EXECUTIVE SUMMARY

This paper assesses the ethical risks of using non-traditional data sources to inform migration-related policymaking and suggests practical safeguards for various stages during the data cycle.

The past decade has witnessed the rapid growth of non-traditional data (social media, mobile phones, satellite data, bank records, etc.) and their use in migration research and policy. While these data sources may be tempting and shed light on main migration trends, ensuring the ethical and responsible use of big data at every stage of migration research and policymaking is complex.

The recognition of the potential of new data sources for migration policy has grown exponentially in recent years. Data innovation is one of the cross-

cutting priorities of IOM's Migration Data Strategy. Further, the UN General Assembly recognises rapid technological developments and their potential in achieving the Sustainable Development Goals and the Global Compact for Safe, Orderly and Regular Migration highlights the importance of harnessing data innovation to improve data and evidence for informed policies on migration.

However, with big data comes big risks. New technological developments have opened new challenges, particularly, concerning data protection, individual privacy, human security, and fundamental rights. These risks can be greater for certain migrant and displaced groups. The identified risks are:



BIG DATA IDENTIFIED RISKS



BIAS AND DISCRIMINATION

Though big data presents new ways of sampling populations, even with a larger sample size, the data collected may not be representative.



DATA PRIVACY, PROTECTION, AND SENSITIVITY

Public and private stakeholders may have profound differences in understanding the risks at stake.



ACCOUNTABILITY & TRANSPARENCY

Differences between the public and private sectors may lead to reduced transparency and accountability, such as varying incentive and risk structures in the management of algorithm-based automated data collection and analyses.



INFORMED CONSENT

Big data tend to be re-used for multiple research objectives, of which not all are always anticipated when requesting and receiving informed consent from data subjects.



DATA SECURITY

Despite security precautions, big data are particularly targeted by cyberattacks which may lead to large scale data breaches.

To reconcile the need for big data and new technological developments to shed light on migration trends and research with the needs for ethical and data protection considerations, many stakeholders have developed normative instruments such as the EU's General Data Protection Regulation (GDPR), the IOM Data Protection Principles (IOM, 2010), or ICRC Handbook on Data Protection in Humanitarian Action (2020). Suggested safeguards should be considered throughout the research project cycle, discussed and agreed upon with research partners:

RECOMMENDATIONS

- » *Assess the suitability of private sector data to serve their specific research interests and approach the private sector with a clear ask, purpose, tools and indicators.*
- » *Due diligence check of all involved partners and clarify stakeholders' involvements (partnerships agreement, organisational values review, and current data security practices).*
- » *Risk-benefit analysis* conducting a Data Protection Impact Assessment (DPIA).
- » *Policies for using or re-using data for research* and how this can be differentiated from commercial research objectives.
- » *Defined Strategies to gain informed consent of data subjects.*
- » *Framework for the partners' roles and responsibilities.*
- » *Policies on data storage, sharing, and retention.*

1. INTRODUCTION

In the information age, more data are produced than ever before. Over the past years, their volume nearly doubled from an estimated 33 zettabytes in 2018 to 71 zettabytes in 2021 – equal to the compound storage of 71 billion 1TB laptops. Until 2025, this amount is expected to grow to a total of 180 zettabytes (Statista, 2021). The exponential increase in data availability and processing capabilities has introduced an array of challenges and opportunities across many policy domains, including migration. The complexity surrounding internal and cross-border human mobility has pointed out the need for reliable, timely, and detailed information to design and assess the impact of migration policies and programmes – a need that to date is only

partly met by traditional data sources, such as household surveys, population censuses and administrative records.

Over the past few years, the rapidly growing use of new data sources in the field of migration, including social media, mobile phone and satellite data, as well as innovative methodologies, such as those based on artificial intelligence (AI) and machine learning, has helped to enhance the understanding of migration-related topics at the local, national, and international levels (Zagheni et al., 2012; Flowminder, 2021; Pötzschke et al., 2017; Rowe et al., forthcoming; Spyrtos et al., 2019; Müller et al., 2020; Milano, 2017; Böhme et al., 2020). Several efforts are underway to develop new guidelines to mitigate the risks and exploit the potential of "big data" for policymaking in various fields, including migration and human mobility (UNESCO, 2021; Franklinos et al., 2020; Floridi, 2018). Ensuring the ethical and responsible use of big data at every stage of migration research and policymaking will be critical for this work to benefit countries and migrants alike.

“

(...) The rapidly growing use of new data sources in the field of migration (...) has helped to enhance the understanding of migration-related topics

This paper seeks to assess the ethical risks of using non-traditional data sources and methods in migration, specifically from a policy perspective. In particular, it outlines how risks of re-using private sector data for migration policy are (or are not) adequately covered by current normative standards and practices in this field, and explore central elements required in future frameworks to ensure their ethical and responsible use. Based on this analysis, the paper concludes by underlining the need to develop further guidance in this field, in particular, due to the various risk implications for vulnerable individuals and groups as a consequence of migration-related research using "big data".

2. THE POTENTIAL OF USING INNOVATIVE DATA SOURCES TO INFORM MIGRATION POLICY

Innovative data sources such as social media, mobile phone data, and satellites offer vast amounts of data relevant for migration governance and research. They can complement traditional data sources and help overcome common challenges, such as data gaps, high costs of acquisition, time lag between data collection and availability of results, limited detail, and coverage of certain forms of human movements – especially of seasonal and temporary nature – and hard-to-reach populations. Through the Data Innovation Directory (Migration Data Portal, 2021), the International Organization for Migration (IOM) has collated information on more than 60 projects using data innovation in the realms of migration and human mobility, such as measuring internal migration through mobile phone data (Lai et al., 2019), estimating the stock of international migrants with social media data (Spyratos et al., 2019), or mapping forcibly displaced populations through satellite data (Bharti et al., 2018).

The COVID-19 pandemic has strengthened the case for data innovation by severely affecting countries' capacities to collect data from traditional data sources and by increasing the appeal of re-using already available data owned by various private sector entities that may be relevant for the analysis of migration-related phenomena. For example, in the Business-to-Government initiative between the European Commission and European mobile network operators (MNOs), aggregated and anonymized phone location data from 14 MNOs across 19 EU Member States have helped understand how migrant communities in different countries have been affected by the pandemic and inform appropriate measures to counter this disproportionate impact (Iacus et al., 2020; Santamaria Serna et al., 2020). Similar initiatives emerged on the local level; for instance, the City of New York established the NYC Recovery Data Partnership with numerous partner organisations

in the private and public sectors to aid the government in COVID-19 response and recovery efforts (NYC Analytics, 2020). These and other efforts are indicative of the potential of new data partnerships between stakeholders in the private and public sectors for informing migration policy and research.

The recognition of the potential of new data sources for migration policy has grown exponentially in recent years. In 2018, the UN General Assembly acknowledged the far-reaching effects of rapid technological developments on achieving the Sustainable Development Goals (SDGs) (UN CEB, 2021). The Global Compact for Safe, Orderly and Regular Migration – the first-ever intergovernmental agreement in the field of migration – highlights the importance of harnessing data innovation to improve data and evidence for informed policies and a balanced public debate on migration. Data innovation is also one of the cross-cutting priorities of IOM's Migration Data Strategy (IOM, 2021), which mentions specific interventions, such as (Ibid: 24):

Investigate new data sources and technologies, such as big data, and pioneer novel methods of data analysis, including methodologies for integrating new data solutions with complementary data sources for quality analysis and dissemination and communication of findings.

And

Promote the development and application of dedicated quality and ethical standards and safeguards for using and sharing such data in accordance with international law.

Data innovation and thematic partnerships also form two strategic pillars of IOM's Internal Displacement Data Strategy 2021-2025 (IOM, 2021). IOM's Global Migration Data Analysis Centre (GMDAC) has implemented several activities within the framework of the Big Data for Migration Alliance (BD4M, 2021) – a joint initiative convened with the European Commission's

Knowledge Centre on Migration and Demography (KCMD) and the GovLab. This collaboration seeks to strengthen the link between the private and public sectors on migration data. It aims to facilitate data partnerships to exchange good practices, build capacities on data innovation across low- and high-income countries, and accelerate the responsible and ethical use of new data sources and innovative methodologies to inform migration policies and programmes. Similarly, the Inter-Agency Working Group on Artificial Intelligence (IAWG-AI) and the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) pursue a multi-stakeholder approach. The IAWG-AI connects experts within the UN system to support the Chief Executives Board (CEB) and High-Level Committee on Programmes (HLCP) on ethics in AI and strategic capacity development (UN CEB, 2021). Further, UN Global Pulse – an initiative of the UN Secretary-General – aims to contribute to “a future in which big data and artificial intelligence are harnessed safely and responsibly for the public good” (UN Global Pulse, 2021) and provides guidance and technical means to governments worldwide. Finally, the IOM Displacement Tracking Matrix (DTM) – a system first implemented in Iraq in 2004 to track and monitor displacement and mobility – has been integrating web scraping and cloud technologies into their data collection and analysis methodologies (World Economic Forum, 2021), while also contributing to ethical frameworks for the responsible application of advanced data science methods for humanitarian outcomes (DSEG, 2020).

3. EARNING PUBLIC TRUST AMID THE RISKS OF RE-USING PRIVATE SECTOR DATA FOR MIGRATION POLICY AND RESEARCH

3.1 Big data, big risks: The circumstances of vulnerable individuals and groups

As the lion's share of novel data is collected and held by organisations in the private sector, the development of cross-sectoral partnerships is critical to unlocking their potential for migration policy. The EU Data Strategy (European Commission, 2020) and the UN Data Strategy (UN, 2020) similarly emphasize data-sharing partnerships as a key enabling factor to tackle and solve policy challenges across several fields, including migration, sustainability, and emergency response. However, much of this potential remains untapped due to a series of challenges, such as the absence of appropriate regulatory frameworks and difficulties in building cross-sectoral partnerships to ensure effective and secure sharing of private datasets.

Novel opportunities have heightened existing risks and given rise to a set of new questions and challenges, particularly in relation to data protection, individual privacy, human security, and fundamental rights (Beduschi, 2020; Molnar, 2019). While these risks are relevant to all population groups, they can be greater for certain migrant and displaced groups, such as individuals displaced by conflict and violence, migrants in vulnerable situations and in need of humanitarian assistance, or those lacking legal status in-country. For instance, unethical data usage can support political persecution (UNHCR, 2021), expose individuals to targeted (false) information (Gibney, 2018), and

exacerbate discrimination (COE, 2018: 25). While such groups may depend on the availability of timely and disaggregated data for their wellbeing, inappropriate or unethical management of such data can also harm them. Amid rapid technology advancements and increasing possibilities of re-using private datasets for various purposes, including providing targeted humanitarian assistance to migrants in need, it becomes vital to ensure that the usage of big data is fully compliant with fundamental ethical and data protection principles at every stage of the data cycle – collection, processing, and dissemination.

Given the novelty and risks related to these approaches, a key overarching challenge is to earn the trust of the public. Whereas national statistical offices in countries around the world have a legal mandate to collect and analyse data on national populations, big data are generally collected by technology companies often based outside users' countries through the mere actions of users of mobile phones or social media, who have varying degrees of awareness of how personal data are used by service providers. This aspect induced many societies to become sceptical and wary towards the idea of sharing information via online platforms, particularly against the backdrop of various high-profile incidents of violated data protection, security, and privacy standards by private entities. Key questions here are how to regulate the re-use of private data for policy in ways that not only respect individuals' privacy but that do not – advertently or inadvertently – harm the individuals that governments and other actors are supposed to serve.

3.2 Risk implications of re-using private sector data under conventional regulatory principles

The risks under the first conventional regulatory principle of data privacy, protection, and sensitivity emerge mainly from two fields of tension that may enable surveillance, persecution, and undermining human rights (DSEG, 2020; ICRC, 2020). First, stakeholders in public and private sectors may have profound cross-sectoral differences in understanding the risks at stake. For instance, while public and non-profit sectors typically place emphasis on the research subjects' data privacy and protection, private sector companies may view proprietary information as most sensitive as it is inextricably linked to maintaining their revenues and organisational wellbeing (OCHA, 2020).

The second field of tension is between the intended and the unintended objectives of practitioners and researchers of re-using private sector data in migration policy. In the case of traditional data collection, it is an established good practice that researchers and practitioners minimise the collected data and specify their intentions for their use as much as possible to (only) achieve a narrowly defined purpose. Yet, this concept runs counter to collecting and analysing big data, which are generated automatically through the various digital services of private sector companies and include highly granular information (ICRC, 2020; UNSDG, 2017). Moreover, some proponents of big data argue that their true value lies specifically in their re-use for originally unforeseen purposes (Cukier and Mayer-Schonberger, 2013). The challenge of maintaining purpose specification might be further heightened as



(...) much of this potential remains untapped due to a series of challenges, such as the absence of appropriate regulatory frameworks and difficulties in building cross-sectoral partnerships to ensure effective and secure sharing of private datasets.



practitioners and researchers could request and receive more detailed data from the private sector than required due to relatively little experience working with big data in migration policy (OCHA, 2020). Also, analyses based on re-using private sector data can be redeployed in a different context than migration practitioners and researchers intended. For instance, the same techniques used to analyse migration patterns could be applied to inform strategic border closures and undermine the right to territorial asylum – instead of providing humanitarian support to vulnerable migrants (DSEG, 2020). This risk is exacerbated by context-based data sensitivity; for instance, the locations of hospitals or participants' religious identities are considered non-sensitive in regular national censuses, but this information can quickly become sensitive in situations of conflict, persecution and displacement (DSEG, 2020). Furthermore, the outputs of big data analytics can also introduce new sensitivities, particularly when datasets are combined, and techniques such as data mosaicking may allow for the reidentification of previously anonymised data (ICRC, 2020). Thus, when evaluating the risks of their big

data projects, researchers should be mindful of what practices their projects could enable in the future.

While informed consent is largely oriented around individual rights, ethical dilemmas relating to bias and discrimination draw attention to the potential of group-level harm. Because of the large pool of subjects available through big data, there is a temptation to view these subjects as the population, rather than a sample, particularly since the large N of big data projects may reduce the importance of common 'small data' concerns, such as statistical significance (Cukier and Mayer-Schonberger, 2013). However, even with large sample size, the data collected from smartphones and computers may not be representative, as access to such devices may differ along gender, age and income lines (DSEG, 2020). Additionally, to the extent that big data reflect existing and historical patterns of discrimination within society, algorithms and AI relying on these data will produce biased results (Leslie, 2019). The problem of discrimination is further exacerbated when algorithms themselves are designed in biased ways, in a phenomenon described as "hard-coded bias" (Leslie 2019). More broadly, communities lacking power may be more likely to encounter automated decision-making and experimental technologies (O'Neil, 2017). Given the opaqueness of algorithms, it is often difficult to challenge the decisions they reach, particularly where no recourse to a human decision-maker is provided (Leslie, 2019). As a result, vulnerable migrant populations may be denied entry into a country, legal status or humanitarian aid on the basis of discriminatory data, coding, application of experimental technologies and access to human decision-makers.

Third, in the realms of accountability & transparency, re-using private sector data in migration policy hosts a series of technical and non-technical risks. First, on the former, similar to research using traditional data, the analyses of private sector data can be erroneous, arbitrary, or simply untransparent (Leslie, 2019). Yet, innovative automated methods increasingly deployed in the private sector to collect and analyse data can intensify these risks by reducing transparency during data lifecycles, which may complicate the identification of parties to hold accountable if individuals or groups are

“

(...) All data must be collected by lawful and fair means; purposes for which personal data are collected and processed must be specified, legitimate, and shared with the data subjects at the time of data collection.



harmed as a consequence of a research project (Leslie, 2019).

Further, the main non-technical risks emerge due to the different incentives of stakeholders in the private and public sectors. In the private sector, algorithms are considered proprietary information (OCHA, 2020), whereas migration policy practitioners and researchers insist on the legibility of algorithms and the need to provide clear explanations of findings to partners and data subjects (DESG, 2020). Further, the different incentive structures can lead to conflicts long after the data is collected and analysed; for instance, for-profit organisations could leverage the results for monetary objectives, such as profiling of migrants for insurance companies or banks assessing the premium rates of individuals with a migration background (DESG, 2020; ICRC, 2020). Consequently, these differences may negatively affect both the partnership and migration policy practitioners and researchers, particularly if those engagements affected vulnerable groups (OCHA, 2020). As a result, partners in private and public sectors may place different emphases on the importance of publishing data and algorithms, which may lead to varying practices and standards regarding the accountability and transparency of re-using private sector data in migration policy (UNSDG, 2017).

The protection risks introduced by data collection reinforces the importance of obtaining informed consent from data subjects. Under the small data framework, an individual provides permission to process their data only for the purposes specified when consent was obtained (ICRC, 2020). However, the sheer size of big data and their granularity create constant opportunities for repurposing that were not anticipated at the time of collection (ICRC, 2020). Moreover, depending on the modes of data collection and the format in which researchers receive the data, it may not be possible to contact data subjects to obtain consent. While consent is not the only legal grounds under which data can be processed, particularly when migration researcher organizations can assert a vital, public or legitimate interest, the re-use of previously collected data for new purposes can result in new protection risks of which data subjects receive little to no notice or information (ICRC, 2020; UNSDG, 2017). Additionally, the temptation to avoid legal liabilities

relating to informed consent may encourage migration researchers to select contexts lacking data regulation (DESG, 2020). As a result, these environments become fertile grounds for testing experimental technologies, with heightened risks of harm for the populations living there. Even where consent can be obtained for re-use of data, its validity is undermined when access to services or products is conditioned on the provision of data or when subjects lack the data literacy to understand the implications of their consent (DESG, 2020; UNSDG, 2020). In sum, data subjects might not be aware that innovative methods were at play, and this lack of awareness can potentially erode other data rights, such as the right to opt-out of data collection, view one's data, correct errors, or lodge appeals (DESG, 2020).

Fifth, just as for traditional data, ensuring data security is a further challenge. Typically, companies collecting and employing big data possess the resources necessary to develop secure data infrastructures and expertise to maintain good practices. Yet, security precautions may vary, and breaches and leaks can certainly occur in the private sector, just as in every other sector (OCHA, 2020). Additionally, sharing data can introduce more opportunities for data breaches. Re-using private data for migration policy and research may encourage belligerent actors to take advantage of the weakest link, as observed with non-governmental organisations operating in insecure settings (DESG, 2020). Further, technologies used for sharing can heighten the data security risks, particularly when data are accessible on a cloud (UNSDG, 2017).

The failure to develop common data security standards for re-using private sector data in migration research and policy would expose both organisations and migrants to severe risks. First, unsecured data can harm migrants, particularly if the breached data holds information on the location and identity of persecuted individuals (OCHA, 2020; ICRC, 2020). For governments, companies, and NGOs, breached data would mean serious reputational risks that can lead to the loss of public trust and subsequently undermine a wide scope of their (not-)for-profit operations.

4. CONTOURS OF FUTURE GOOD PRACTICES AND REGULATORY FRAMEWORKS ENSURING THE ETHICAL USAGE OF PRIVATE SECTOR DATA IN MIGRATION POLICY

4.1 Contributions of current standards and practices for developing ethical private-public data partnerships

Established normative instruments, such as the EU's General Data Protection Regulation (GDPR), are a substantive start to adapt digital safeguards and ethical standards to protect individuals in the digital space. Additionally, a range of non-normative standards is contributing to ensuring the ethical and responsible use of private sector data in migration-related research. Several handbooks and tools developed for the humanitarian and development sectors offer a useful starting point for the conceptualisation of an ethical framework geared towards the re-use of private data for migration research and policy. For migration-related specifics, the International Organization of Migration (IOM) has developed its IOM Data Protection Principles (IOM, 2010). Beyond migration, the UN Sustainable Development Group (UNSDG) crafted a guidance note applying data protection principles to big data analytics within the development sector (UNSDG, 2017). Additionally, a chapter of the International Committee of the Red Cross's (ICRC) Handbook on Data Protection in Humanitarian Action translates data protection principles to the big data realm, given the numerous tensions between the organizing principles of big data analytics and

data protection principles designed for small data collection (ICRC, 2020). More specifically, the UN Office for the Coordination of Humanitarian Affairs' (OCHA) Centre for Humanitarian Data released a guidance note addressing partnerships between public humanitarian organizations and the private sector, considering the potential of the advanced information and communications technology (ICT) products and big data sources available within the private sector (OCHA, 2020).

These and other ethical standards highlight fundamental aspects of ethical data use for policy-related research projects. Namely, all data must be collected by lawful and fair means; purposes for which personal data are collected and processed must be specified, legitimate, and shared with the data subjects at the time of data collection. At the same time, existing international frameworks cannot sufficiently address all (potential) harm caused by the use of big data, particularly concerning algorithm-based decision-making (McGregor et al., 2019). The recent elaboration of a Recommendation on the Ethics of AI (UNESCO, 2021) from a multi-stakeholder group is reacting to this concern and builds upon a preliminary study of the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), which defined a set of key principles relevant for future norm-setting, including accountability, transparency, inclusiveness, democracy, and human rights (COMEST, 2019: 24). The European Commission proposal for a regulation of AI (or 'AI Act') published in early 2021 is a welcome and essential step in this direction (European Commission, 2021). Further, specifically for the application of AI methodologies, the Alan Turing Institute's guide to AI ethics provides a framework for identifying opportunities for bias and discrimination at various points along the project development process, as well as broader societal implications stemming from the use of AI (Leslie, 2019).

The following section articulates potential strategies that can be used to alleviate the outlined risks and navigate cross-sectoral partnerships involving the re-use of private sector data as relates to migration.

4.2 Elements of an ethical framework for the re- use of private data to inform migration-related policy

Rather than offering a set of prescriptions, the following recommended elements emphasise a process-driven approach that considers the purpose of research, the level of aggregation and sensitivity of the data, the new risks such data analysis could introduce, and the presentation of findings. While certain ethical standards must form a baseline for all research projects, the wide variety of purposes for which migration research is conducted reinforces the need to account for the specificities of each project. For example, migration research conducted for the purposes of informing humanitarian response may require a higher level of granularity to be operationally relevant than a research project focused on broader trends over time and space. In turn, these more sensitive data would necessitate stronger data security measures and more restricted publication of findings.

As a first step, migration researchers and practitioners should assess the suitability of private sector data to serve their specific research interests. Researchers should consider the scale and duration of the project, potential risks, and past projects conducted by other organizations that could offer useful insights on data management and best practices (DSEG, 2020). Having determined that big data analytics are well-suited for the research question, migration practitioners and researchers should approach potential partners in the private sector with a clear articulation of their research purpose and tools and indicators that could help them meet their aims. At this stage, private organizations would then examine how their existing data and technologies could fill the mentioned research gap with the goal of minimising the volume and detail of datasets necessary to answer the research question. In this sense, big data provided to migration practitioners and researchers by private companies should be fit for purpose, even if the original dataset contains information extending beyond the scope of the research project.

Second, once suitable partners have been identified based on the data needs, a due diligence check of all involved partners should follow. As the priority concerns for each organization may vary, this process can account for a broad range of criteria, such as past partnerships, organisational values, and current data security practices (OCHA, 2020). Subsequent communication of the due diligence procedures to the public can further improve perceptions of the partnership and project (OCHA, 2020).

Next, partners should conduct a risk-benefit analysis, for instance, by conducting a Data Protection Impact Assessment (DPIA), which provides an opportunity to define sensitive data within the project's context and anticipate various potential risks, such as data breaches, leaks, data access by third parties, and unclear responsibility distribution among partners. The DPIA can also support outlining appropriate policies for these potential scenarios and contribute to establishing mutually agreed data protection measures (ICRC, 2020; OCHA, 2020). Additionally, as data protection risks are often context-specific, the researchers should consult with members of the population under study to better understand the distinct risks the group may face. This exchange between partners from diverse sectors and backgrounds will also help identify sources of bias at various stages and uncover potential blind spots (DSEG, 2020).

As a fourth element, the partners should define policies for the use of results of the completed research project, such as retention and deletion of data and organisation-specific concerns (DSEG, 2020). For example, research organisations may seek to limit the extent to which data isolated for research purposes can be subsequently used for commercial objectives, while private organisations may move to restrict future use of the data, findings and technologies as a matter of intellectual property (ICRC, 2020; OCHA 2020). Further, if shared in an anonymized form, partners should also weigh the possibility of reidentification through consideration not only of the dataset being examined but also potential datasets that could be used for matching (ICRC 2020). Depending on the project objectives,

aggregation, pseudonymisation or techniques such as data smoothing can be employed to reduce the sensitivity of the dataset (Welsh et al., 2021; ICRC, 2020). Lastly, to prevent external users of the project results from eroding data subjects' rights, partners should devise a strategy to vet third parties seeking to access the data, accounting for data security practices, their human rights track record, and other areas of concern (DSEG, 2020).

The following element suggests pursuing creative strategies to gain informed consent of data subjects. In cases where contact with data subjects is possible through the private company, said company could devise a text message or pop-up window informing data subjects of the project and obtaining their consent (ICRC, 2020). These communication channels can also be used to publicize the subjects' right to correct or erase their data or file a complaint (ICRC, 2020). If it is not possible to contact individual subjects directly, partners should discuss other methods for informing subjects of the project, such as advertisements on highly frequented websites or notices in printed communication materials such as newspapers or posters (ICRC, 2020). To reduce the risk of harm to subjects, partners should also investigate strategies that balance the level of detail needed for research against subjects' safety through techniques such as anonymization, aggregation, and other advanced techniques.

Closely interconnected with the previous element, the partners should define a framework for the partners' roles and responsibilities to ensure and improve accountability and transparency of their research project using private sector data relevant to migration policy (DSEG, 2020). These frameworks can also address issues related to error and failure. For example, in situations of automated decision-making, partners can establish whether false positives or false negatives would be more harmful to subjects (DSEG, 2020). Depending on the stakes of the project, partners can also agree upon an acceptable level of error in their findings (DSEG, 2020). In the event of system-level failure, partners should determine contingency plans and methods for rectifying any harms experienced by data subjects (DSEG, 2020). For projects involving automated decision-making, researchers should ensure a layer of human review of an algorithm's findings if

the decisions being made could negatively impact the wellbeing and rights of migrants (ICRC, 2020). Furthermore, the partners should come to an agreement regarding the extent to which information on the algorithm responsible for data processing can be publicly discussed, while bearing in mind the potential risks that could stem from oversharing (ICRC, 2020). For example, even if the components and mechanisms of an algorithm must remain secret for proprietary reasons, researchers could consider disclosing whether an algorithm was used and discussing its decision-making processes in general terms. This would enable partners to explain the algorithm in broad and easy-to-understand language while safeguarding the company's intellectual property (UNSDG, 2017).

Finally, the last element suggests that the research team jointly devise policies on data storage, sharing, and retention. These policies serve to maintain the safety and security of data subjects on the one hand and proprietary information of private companies on the other hand (DSEG, 2020). When sharing data, partner organisations should consider encrypting sensitive datasets sent over a server (UNSDG, 2017). De-identification and aggregation of datasets by private companies prior to sharing can also reduce the sensitivity of data (UNSDG, 2017). Additionally, the number of individuals within each organization who can access the data should be reduced to a minimum and trained well to alleviate risks of negligent data management (UNSDG, 2017, DSEG, 2020). Relatedly, organisations should develop procedures to record which individuals access the data and notify partners in the event of unauthorized access (Council of Europe, 2017). With respect to storage, partners should weigh the risks of hosting data on local directories versus shared clouds (UNSDG, 2017).

4.3 A hands-on approach alleviating the risks of using private sector data for migration policy under current frameworks for traditional statistics

Given the tensions between current practices and incentive structures of using big data in migration policy between private and public sector partners, minimum standards need to be developed for the ethical data use and navigation of cross-sectoral partnerships. Existing resources developed for humanitarian and development settings can be repurposed with attention to the specificities of migration research, considering its purpose and the level of aggregation at which migration research organizations operate. For example, the DSEG Decision Tree, originally developed for humanitarian practitioners seeking to implement data science projects responsibly, would prove useful to migration researchers as well, as it encourages consideration of ethics at each stage of a project, from conceptualization to communication of findings (DSEG, 2020). The principles underlying the Decision Tree are drawn from broader humanitarian and human rights considerations and ethical guidelines developed specifically for advanced data science methods (DSEG, 2020). Additionally, as an interactive process posing a series of questions regarding research methods to users, the Decision Tree provides further guidance on topics such as data storage and testing and evaluation methods (DSEG, 2020). Here, the DSEG Decision Tree is particularly helpful, as it walks users through questions related to problem recognition and the appropriateness of advanced data science methods for the problem being addressed (DSEG, 2020).

5. CONCLUSION

In the information age, more data are produced than ever before. This paper has assessed the ethical risks of using non-traditional data sources for migration analysis, policy and programming. Specifically, it has outlined key implications of re-using private sector data for migration research projects in conventional areas of ethical risks during the data lifecycle – data privacy, transparency, and informed consent – and drawn the contours of a framework to ensure their ethical and responsible use.

As a starting point, the paper stressed the benefits of using “big data” for migration research, of which a great share are collected and processed by private sector organisations. We described numerous of their attributes complementing traditional data sources in migration policy and research, which can be summarised in three main features: First, big data can provide near real-time observations; second, they offer a fine-grained spatial resolution; and third, they may cover entire populations of users of certain devices and digital platforms.

“

(...) the use of private sector data for research purposes will most likely expand in the future, and we are now at a critical point in time to establish reliable frameworks

Considering the need to improve evidence for informed policymaking in migration and beyond, we argue that the use of private sector data for research purposes will most likely expand in the future, and we are now at a critical point in time to establish reliable frameworks to ensure the ethical and responsible use of these innovative data .

We set out the contours of future frameworks by outlining key elements to ensure the ethical and responsible re-use of private sector data. Existing normative and non-normative standards relevant for safeguarding individuals' and groups' data protection rights in migration-related research pointed to the need to update and adapt current frameworks to accommodate the new risks implied in using innovative data sources and methods in comparison to traditional statistical approaches.

Private sector organisations often have different – at times competing – incentives, particularly relating to issues of transparency and accountability due to issues resulting from regarding algorithms as proprietary information. Additionally, the re-use of existing data for new purposes may require the data subjects' informed consent; yet it may be impossible for researchers to contact subjects, depending on the level of detail in the dataset. Finally, differing data security standards can expose all members of a research partnership to security problems. In view of these and further potential risks, conducting migration-related research with “big data” can be considered a widely unregulated space at the time of the publication of this piece.

The necessary elements we introduced to ensure the ethical use of private sector data can provide guidance for both the design of new research projects and the development of future normative and non-normative standards for secure data governance mechanisms. They comprise seven areas and accompany different stages of the research project and data lifecycle. They seek to apply the logic and precautions of established standards for working with traditional data sources to the challenges and characteristics of working with “big data”. Considering their vast volume and velocity, these elements underline critical steps, such as assessing the necessity to

use big data for the specific research purpose, conducting a due diligence check of all partners involved and a risk-benefit analysis to define the sensitivity of the data and anticipate various project-specific risks. We underlined the need to define policies regulating the use of the results following the completed project and an accountability framework for the partners' roles and responsibilities. Further, we stressed the need to develop creative strategies to gain informed consent of data subjects and to devise secure policies on data storage, sharing, and retention, including data access for third parties.

REFERENCES

Beduschi, A.

2020 International migration management in the age of artificial intelligence. *Migration Studies*. doi: [10.1093/migration/mnaa003](https://doi.org/10.1093/migration/mnaa003).

Bharti, N., Tatem, A.

2018 Fluctuations in anthropogenic nighttime lights from satellite imagery for five cities in Niger and Nigeria. <https://doi.org/10.1038/sdata.2018.256>.

Böhme, M. H., Gröger, A., Stöhr, T.

2020 Searching for a better life: Predicting international migration with online search keywords. www.sciencedirect.com/science/article/abs/pii/S0304387819304900.

Big Data for Migration Alliance [BD4M]

2021 Designing data collaboratives to better understand human mobility and migration in West Africa. <https://odpl.thegovlab.com/uploads/ccm-directus/originals/aa45abf2-743d-4d58-8471-e25bff956ba5.pdf>.

Council of Europe

2018 Discrimination, artificial intelligence, and algorithmic decision-making. <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

Council of Europe

2017 Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data. <https://rm.coe.int/16806ebe7a>

Cukier, K., Mayer-Schonberger, V.

2013 *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray.

Dignum, V.

2018 "Ethics in artificial intelligence: Introduction to the special issue", in: *Ethics and Information Technology*, 20(1), 1–3.

DSEG

2020 A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector. www.hum-dseg.org/sites/default/files/2020-10/Framework%20for%20the%20ethical%20use.pdf

European Commission

2021 Proposal for a Regulation on a European approach for Artificial Intelligence. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

European Commission

2020 A European Strategy for Data. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>.

Facebook

2021 High-Resolution Density Maps. https://data.humdata.org/organization/facebook?sort=metadata_modified+desc.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Schafer, B.

2018 AI4People: An ethical framework for a good AI society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707.

Flowminder Foundation, Ghana Statistical Services, Vodafone Foundation

2021 Mobility analysis to support the Government of Ghana in responding to the COVID-19 outbreak. https://statsghana.gov.gh/COVID-19%20mobility%20resport%20for%20Ghana_3.pdf.

Franklinos, L., Parrish, R., Burns, R., et al.

2020 Key opportunities and challenges for the use of big data in migration research and policy. UCL Open: Environment Preprint. DOI: 10.14323/111.444/000042.v1

Gibney, E.

2018 The scant science behind Cambridge Analytica's controversial marketing techniques. www.nature.com/articles/d41586-018-03880-4.

Google

2021 COVID-19 Community Mobility reports. www.google.com/covid19/mobility/. Iacus, S., Santamaria Serna, C., Sermi, F., Spyrtatos, S., Tarchi, D. and Vespe, M. (2020) How human mobility explains the initial spread of COVID-19. Luxembourg: Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC121300>.

ICRC

2020 Handbook on Data Protection in Humanitarian Action. www.icrc.org/en/data-protection-humanitarian-action-handbook

IOM

2021 Internal Displacement Data Strategy 2021-2025: Strengthening Capacity and Leadership in Internal Displacement Data. <https://publications.iom.int/books/internal-displacement-data-strategy-2021-2025>

IOM

2021 IOM Migration Data Strategy: Informing policy and action on migration, mobility and displacement. <https://publications.iom.int/books/iom-migration-data-strategy-informing-policy-and-action-migration-mobility-and-displacement>.

IOM

2010 IOM Data Protection Manual. https://publications.iom.int/system/files/pdf/iomdataprotection_web.pdf.

Lai, S., Zu Erbach-Schoenberg, E. DP., Pezzulo, C., Ruktanonchai, N., Sorichetta, A., Steele, J., Li, T., Dooley, C., & Tatem, A.

2019 "Exploring the use of mobile phone data for national migration statistics", in: *Palgrave Communications*, 5, [34]. <https://doi.org/10.1057/s41599-019-0242-9>.

Leslie, D.

2019 Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>

Migration Data Portal

2021 The Data Innovation Directory. <https://migrationdataportal.org/data-innovation>.

Milano, L.

2017 Idetect: How technology and collaboration between innovators can help ensure “no one is left behind”. www.internal-displacement.org/expert-opinion/idetect-how-technology-and-collaboration-between-innovators-can-help-ensure-no-one.

McGregor L., Murray D., Ng V.

2019 “International Human Rights Law as a Framework for Algorithmic Accountability”, in: International and Comparative Law Quarterly, 68(2): 309–43.

Molnar P

2019 Technology on the margins: AI and global migration management from a human rights perspective. Cambridge International Law Journal, 8(2): 305-330.

Müller, K., Schwarz, C.

2020 Fanning the Flames of Hate: Social media and hate crime. <https://academic.oup.com/jeea/advance-article-abstract/doi/10.1093/jeea/jvaa045/5917396?redirectedFrom=fulltext>.

NYC Analytics

2020 NYC Recovery Data Partnership. www1.nyc.gov/site/analytics/initiatives/recovery-data-partnership.page.

OCHA

2020 Note #3: Data Responsibility in Public-Private Partnerships. <https://centre.humdata.org/guidance-note-data-responsibility-in-public-private-partnerships/>.

OECD

2020 Good Practice Principles for Data Ethics in the Private Sector. www.oecd.org/gov/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.pdf.

O’Neil, C.

2017 Weapons of math destruction. Penguin Books.

Power, D., Thom, M., Gray, J., Albert, M., Delaporte, S., Li, T., Harrison, J., Greenhalgh, J., Thorne, N., Bengtsson, L.

2020 FlowKit: Unlocking the power of mobile data for humanitarian and development purposes. https://digitalimpactalliance.org/wp-content/uploads/2019/02/FlowKit_UnlockingthePowerofMobileData.pdf.

Pöttschke, S., & Braun, M.

2017 "Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries", in: *Social Science Computer Review*, 35(5), 633-653.

Rowe F., Mahony, M., Graells-Garrido, E., Rango, M., Sievers, N., forthcoming

2021 Using Twitter to Track Immigration Sentiment During Early Stages of the COVID-19 Pandemic in: *Data & Policy*, Cambridge University Press.

Santamaria Serna, C., Sermi, F., Spyrtatos, S., Iacus, S., Annunziato, A., Tarchi, D. and Vespe, M.,

2020 Measuring the Impact of COVID-19 Confinement Measures on Human Mobility using Mobile Positioning Data, EUR 30290 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-20428-2, doi:10.2760/913067, JRC121298.

Spyrtatos S., Vespe M., Natale F., Weber I., Zagheni E., Rango M.

2019 Quantifying international human mobility patterns using Facebook Network data. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0224134>.

UN

2020 UN Data Strategy. www.un.org/en/content/datastrategy/images/pdf/UN_SG_Data-Strategy.pdf

UN CEB 2018

Artificial Intelligence. <https://unsceb.org/topics/artificial-intelligence>.

UNSDG

2017 Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda. <https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda>

UNESCO

2021 Elaboration of a Recommendation on the ethics of artificial intelligence. <https://en.unesco.org/artificial-intelligence/ethics>.

UNHCR

2021 News comment: Statement on refugee registration and data collection in Bangladesh. www.unhcr.org/news/press/2021/6/60c85a7b4/news-comment-statement-refugee-registration-data-collection-bangladesh.html.

UN Global Pulse

2021 About UN Global Pulse. <https://www.unglobalpulse.org/about/>.

Welsch, W., Pandey, B., Rekabsaz, N., Giannone, C.

2021 The Innovation Project: Can Advanced Data Science Methods be a Game-Changer for Data Sharing? www.jips.org/jips-publication/the-innovation-project-can-advanced-data-science-methods-be-a-game-changer-for-data-sharing/.

World Commission on the Ethics of Scientific Knowledge and Technology [COMEST]

2019 Preliminary study on the ethics of artificial intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>.

World Economic Forum

2021 How cloud technology is helping migrants stay safe during the pandemic. www.weforum.org/agenda/2021/04/how-cloud-technology-is-helping-migrants-stay-safe-during-the-pandemic/?fbclid=IwAR2P-gZJN0FuQ-3lsZQSOv5VuSpGyh_FianTSzFJLs_CXmnMrE6C6onXCus.

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).



International Organization for Migration
Global Migration Data Analysis Centre
Taubenstr. 20-22, D- 10117 Berlin, Germany

Tel.: +49 30 278 778 21 | Fax: +49 30 278 778 98

gmdac@iom.int | [@IOM_GMDAC](https://twitter.com/IOM_GMDAC)