TRANSLATIONAL ARTICLE

Challenges and opportunities in accessing mobile phone data for COVID-19 response in developing countries

Sveta Milusheva[1]* , Anat Lewin[2], Tania Begazo Gomez[2], Dunstan Matekenya[3] and Kyla Reid[4]

[1]Development Impact Evaluation Department, World Bank, Washington, District of Columbia, USA
[2]Digital Development Global Practice, World Bank, Washington, District of Columbia, USA
[3]Development Data Group, World Bank, Washington, District of Columbia, USA
[4]Independent Consultant Toronto, ON, Canada
*Corresponding author. E-mail: smilusheva@worldbank.org

Abstract

Anonymous and aggregated statistics derived from mobile phone data have proven efficacy as a proxy for human mobility in international development work and as inputs to epidemiological modeling of the spread of infectious diseases such as COVID-19. Despite the widely accepted promise of such data for better development outcomes, challenges persist in their systematic use across countries. This is not only the case for steady-state development use cases such as in the transport or urban development sectors, but also for sudden-onset emergencies such as epidemics in the health sector or natural disasters in the environment sector. This article documents an effort to gain systematized access to and use of anonymized, aggregated mobile phone data across 41 countries, leading to fruitful collaborations in nine developing countries over the course of one year. The research identifies recurring roadblocks and replicable successes, offers lessons learned, and calls for a bold vision for future successes. An emerging model for a future that enables steady-state access to insights derived from mobile big data - such that they are available over time for development use cases - will require investments in coalition building across multiple stakeholders, including local researchers and organizations, awareness raising of various key players, demand generation and capacity building, creation and adoption of standards to facilitate access to data and their ethical use, an enabling regulatory environment and long-term financing schemes to fund these activities.

Policy Significance Statement

A history of one-off pilot projects, an absence of scalable and sustainable business models, a lack of standards as well as regulatory restrictions have meant that in many cases, aggregated and anonymized mobility data are challenging to access for development projects, if they are made available at all. The research team makes one of the first large-scale efforts to access mobile phone data spanning 41 countries across multiple continents, with a focus on Africa. Nine country cases became fully operational within one year. Lessons are drawn from the experience on the key challenges, successes, and steps needed to inform future, scaled up, and replicated efforts.

1. Introduction

As countries, communities, and individuals around the world have grappled with the unprecedented challenges presented by COVID-19, governments, public health experts, and the development

community have urgently sought innovative ways to respond to the pandemic. Key drivers for understanding the transmission of COVID-19 are population mobility, density, and behavior. Big data, in particular anonymized, aggregated data from mobile phones (call detail record [CDR]-derived indicators)[1] have proven efficacy as a proxy for human mobility and an input to epidemiological modeling (Bengtsson et al., 2011; Wesolowski et al., 2012b; Bengtsson et al., 2015; Wesolowski et al., 2015a; 2015b; Ihantamalala et al., 2018; Milusheva, 2020). Mobility indicators generated from mobile operators' aggregated data can - despite some limitations due to low-level mobile phone penetration in certain geographies or demographic groups - strengthen responses in resource-constrained countries where alternative data sources may not readily be available.

In the context of COVID-19, CDR-derived indicators can offer near real-time insights into patterns of mobility during outbreaks and lockdowns and can follow the impact of public health interventions across the pandemic lifecycle. These data sets, when used effectively and responsibly, offer the potential to rapidly inform more effective policy and operational responses, support improved preparedness, and ultimately deliver better development outcomes. This is particularly important in countries where recent survey data are limited. Additionally, while in some countries mobility statistics have been generated from location data provided by smartphone mobile applications, in many countries, these are not representative of the wider population. This is due to either the limited penetration of smartphones, which for instance is 45% in Sub-Saharan Africa, or to the limited use of location-based applications (GSMA, 2020). The lack of representation is especially stark in the bottom income quintile in Africa, where only 6.6% have access to mobile internet, while 55.3% have access to mobile phones (Frankfurter et al., 2021). In these contexts, CDR-derived indicators can offer more representative measures of population mobility and behavior.

This article builds on the experience of implementing a World Bank initiative to develop standardized, anonymized, and aggregated mobility indicators from CDR data in response to the COVID-19 outbreak and to integrate these into government efforts aimed at locally mitigating the impact of COVID-19. A framework of key elements is developed to facilitate successful use of CDR-derived indicators to inform policies in developing economies and to present different avenues of gaining access to mobile network operator (MNO) data given the characteristics, relationships, and incentives of various stakeholders. Lessons learned are developed from attempts to access CDR-derived indicators in 41 countries, with recurrent roadblocks and success factors identified. Most of the cases (28) are in Africa, where other sources of high-frequency mobility data with high population coverage are limited.[2]

While the premise of mobile phone data's utility is widely acknowledged (Haddad et al., 2014; Blondel et al., 2015), obtaining access to analytics from these data in developing countries remains an uphill battle, as reported in other studies (Matekenya et al., 2021). The challenges have included regulatory restrictions, lengthy processes to obtain government clearances, the need for negotiations on data access agreements, insufficient capacity of data users, high costs, funding gaps, and needs for coordination across actors. The convenience of following a standardized framework for accessing the data in a comparative regional or global format that addresses concerns about the ethical use of these data remains an ambitious goal for

---

[1] CDRs contain information on every call and text, including the IDs of the callers and receivers, the date and time of the interaction and the cell phone tower associated with the call or text. Mobile network operators (MNOs) collect this information for billing purposes, but it can be used to identify the approximate locations from which calls and texts are made by individual SIM cards and measure the cards' movement over time. MNOs can produce pseudonymous versions of these data, where the SIM ID is replaced with a randomly generated ID that still enables identifying the same SIM over time but does not identify the caller or receiver. The data are aggregated into administrative level indicators that can serve as proxies for population mobility and inform related analysis. For the purposes of this article, "CDR-derived indicators" and "mobility data" are used interchangeably. In addition to CDRs, MNOs may also generate signaling data (the passive location updates which devices send to the MNO network regardless of whether the user is interacting with the device or not). Whether the MNO collects these data at all and how long it retains the data if they do collect them varies greatly across MNOs and the technologies they are using. While these data can provide better coverage and representation because they do not rely on phone usage, we found them to generally be unavailable, and therefore the focus has been on CDR-derived indicators.

[2] In countries with higher smartphone penetration, new tools such as Google mobility reports, Facebook mobility data and information from platforms such as Cuebiq and Unacast can provide high-frequency mobility data that can inform COVID-19 modeling and policies.

now. Yet, this initiative led to successful collaborations in nine countries, where the commitment of local and multinational MNOs, government officials, and other parties demonstrated the advances that can be made in leveraging the potential of big data. Creating the necessary conditions to reach this potential more broadly will be slow and will require concerted leadership, coordination, standardization, and governance across multiple stakeholders.

COVID-19 is not the first health crisis for which the potential benefits of CDR data were recognized. At the start of the Ebola outbreak in 2014, Wesolowski et al. (2014) published a commentary on the potential benefits of CDR data to monitor population mobility and outbreaks. Due to data protection regulatory frameworks inconducive to the rapid sharing of anonymized data, use of CDR-derived indicators remained limited during the outbreak, which was seen as a missed opportunity.[3] During the five years between the Ebola outbreak and the onset of the COVID-19 pandemic, the development community and MNOs held extensive dialogue on how to move this agenda forward; however, progress scaled beyond promising pilot projects has been elusive (GSMA, 2018; 2019). When COVID-19 began, developing countries, MNOs and the development community found themselves again in a situation in which a call to action for the use of CDR data was issued (Buckee et al., 2020; Oliver et al., 2020). This article recommends actions to support a model for the future that enables steady-state access to indicators to inform policy making and development interventions. Laying the foundations for this requires preparation, investment in coalition building, awareness raising, demand generation and capacity building, creation and adoption of standards for data access and ethical data use, an enabling regulatory environment and a long-term financing scheme to fund these activities. Such a steady-state model could then define and enable rapid and resource-minimal response at times of health, environmental, or humanitarian crises.

2. Approach

In order to successfully access CDR-derived indicators and translate the data into useful insights, a number of elements need to be in place to ensure a productive consensus among a diverse set of stakeholders (Figure 1). Among these, MNOs and government agencies are key stakeholders for the supply and demand of data, respectively. Data translators (international organizations, nongovernmental organizations, academia, and private sector) can facilitate the relationship between MNOs and government, bring technical expertise on data analytics and policy insights to the table, tailor data sharing agreements and convene consensus among stakeholders. The approach used for this initiative had three main elements. The first was building demand within relevant government agencies for these analytics and identifying local organizations and researchers open to collaboration. The second was building collaborative partnerships with MNOs, government agencies, and other partners to facilitate access to aggregated data. The third was the data analysis, both in terms of supporting and facilitating the creation of the aggregated data set as well as the application of the indicators to use cases in collaboration with government and local research teams.

2.1. Data demand

The data demand side is central to this work. Ministries of Health and their epidemic response teams are key beneficiaries of mobility data insights. In certain countries, COVID-19 task forces were established to coordinate the collection and integration of data that could support management of the crisis. A global health pandemic such as COVID-19 has impacts that reach far beyond health. Social protection agencies and organizations in charge of supporting the well-being of the population are also important actors for whom better data on population mobility and density can help to achieve better program outcomes. Engagement between local and international researchers, data analysts, and stakeholder government agencies is important to understanding the local context and beneficiaries' needs. Therefore, in each country case, locating the demand for these analytics and learning from the relevant government and

---

[3] An eventual evaluation of travel policies was published several years after the outbreak (Peak et al., 2018).
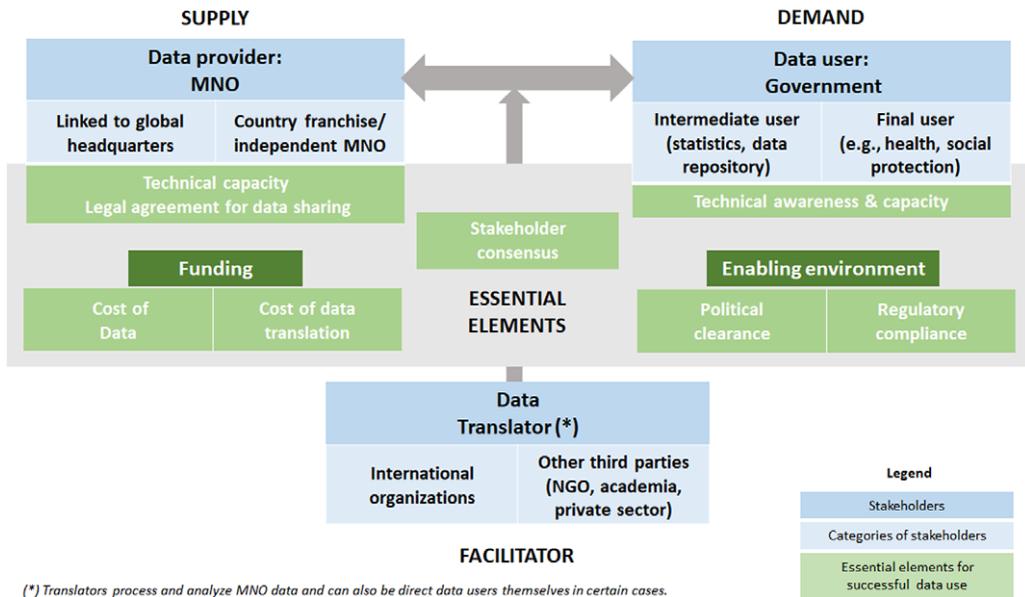
Figure 1. *Stakeholders and key elements for the successful use of CDR-derived indicators.*

researcher counterparts was important to ensure that the data can effectively inform policy action. Local ownership was likewise important for facilitating data access from the MNOs.

2.2. Data access

Based on the country cases pursued, three main approaches emerged for facilitating data access. First, data access agreements can be signed by an interested party directly with an MNO to gain access to the data. Second, data access agreements can be signed by a government agency with the MNO—typically this was the national telecommunications regulatory authority, in select cases with the involvement of a national crisis response task force or a national statistics agency. The interested party could then access the data through the government agency. Finally, data access can also be channeled by the government or a local MNO via third-party organizations (TPO; either for-profit, academic, or nonprofit institutions) who have an agreement with the government or the MNO. The recommended approach depends on a combination of factors, including the statistical and analytical capacities of the government agencies and the effectiveness of their relationship with MNOs, the level of trust and the alignment of expectations among stakeholders. A standard approach was not identified—the path to success was determined by local contexts.

Of the 41 countries in which the World Bank explored moving forward, after evaluating potential sensitivities in each country, 37 cases remained in which data access was pursued. Figure 2 breaks down the main approaches used in those cases.[4]

2.2.1. Working with MNOs

The initiative's work with MNOs followed one of two scenarios: (a) working with the headquarters of a multinational MNO that provides umbrella facilitation for access to data in several countries and (b) working with an independent MNO or an MNO franchise in a country.

Multinational MNO. Working directly with the global headquarters of a multinational MNO can be more efficient when the MNO's structure is centralized; this can facilitate access to data in several

---

[4] Note that in some cases, multiple approaches were tried, but only the main one is counted in the figure.

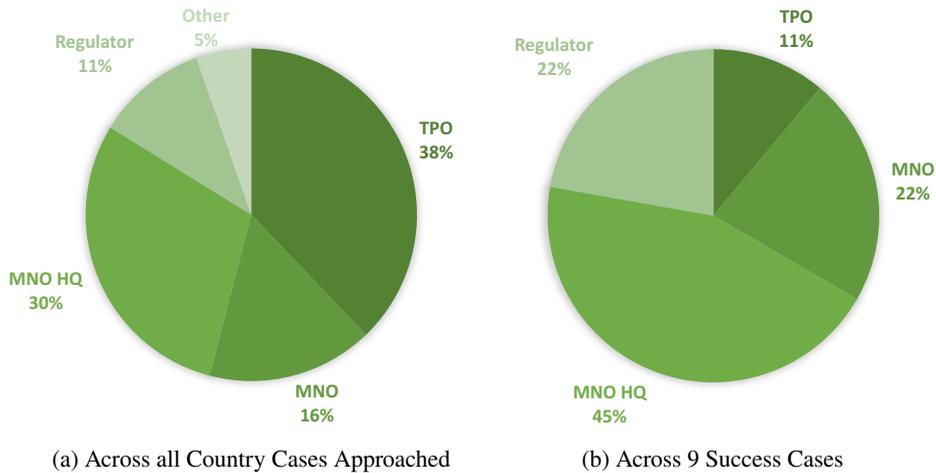(a) Across all Country Cases Approached   (b) Across 9 Success Cases

Figure 2. *Data access approaches. Note: Mobile network operator (MNO) refers to accessing the data via an independent or franchise MNO in-country, while MNO HQ refers to accessing the data via the headquarters of a global MNO. Third-party organization (TPO) refers to universities, for-profit firms and nonprofit organizations.*

countries as part of one process. This was the strategy initially followed to maximize the number of country cases in which data access could be pursued. For instance, working with one MNO, Vodafone, at the central level enabled an agreement that facilitated data access in four country cases. Preparing this data sharing agreement required significant work on the part of both legal teams in order to meet the requirements of both organizations. Nevertheless, once it was signed, the number of country-level access agreements to be negotiated was significantly reduced. However, regulatory and other governmental approvals and concurrences were still needed for compliance with national legislation; therefore, not all four of these cases moved forward.

Working directly with the hub of a multi-national MNO was less effective where the relationships between headquarters and the local MNOs were decentralized. In some cases, a productive relationship with headquarters was helpful to facilitate discussion at the country level; nevertheless, it required establishing a relationship directly with each country-based MNO as well. This produced multiple parallel processes.

Individual MNO. Working directly with individual MNOs at the country level can be more straight-forward, particularly when coordinating with data-experienced MNO staffers. Yet, competing demands faced by MNOs during the COVID-19 crisis of providing broadband and telecom services to a larger population working from home as well as supporting new initiatives such as COVID-19 mobile information campaigns affected the MNOs' bandwidth to establish new relationships or delve into a complicated data project.

An additional challenge arose in several discussions on the availability of funding for the production and usage of the CDR-derived indicators. Due to the limited short-term availability of financing for COVID-19 data initiatives, it was possible to move forward primarily where MNOs were willing to provide the needed analytics pro bono or where these analytics could be part of their corporate social responsibility (CSR) or business strategy. Given the devastating impacts of the COVID-19 pandemic, most of the MNOs that agreed to share data did so as part of their CSR.

Country case study 1. The first country in which CDR-derived indicators were successfully accessed and analyzed to support COVID-19 response is a case study of preparedness: an existing agreement with a local MNO facilitated faster access to aggregated analytics from CDR data from a prior model of a cholera

outbreak. Since the legal and practical agreements were already in place, along with trust among the partners, accessing newer data to support epidemiological modeling for COVID-19 was a straight-forward exercise.

The MNO agreed to share aggregated data based on the prior data sharing agreement, and the needed regulatory approvals were granted based on this existing partnership. The research team wrote code for producing the aggregated indicators for the modeling and disease analytics. The code was packaged in a container via Docker, which allowed the technical team of the MNO to run it on the local MNO's system. This limited the amount of technical work required from the MNO team and served to set up a system in which all sensitive data were processed locally, on the premises of the MNO. Aggregated datasets were shared with the researchers, ensuring user privacy and safe use of the data in accordance with best practices (De Montjoye et al., 2018). While technical challenges arose in this process that took time to resolve, with sign-off and support at the highest level of the MNO, those challenges could be addressed through close collaboration between the researchers and the technical team. The resulting datasets were used to produce insightful analyses that informed health, lockdown, and preventive policy measures.

A lesson learned is to make the necessary investments up front and maximize preparation before a crisis begins, so that during a crisis, the focus can be on execution.

2.2.2. Working with government regulatory authorities

Since telecommunications regulatory authorities issue operating licenses to MNOs in all countries, they are natural partners for facilitating access to mobility data for development work. In some cases, this can mean all MNOs in the country pass CDR-derived data through to the regulator, and the interested party could access the data with permission from the regulator. In other cases, the regulator facilitates and enforces rules for the agreements, paving the way for data to be accessed directly from an MNO. In selected cases, it is the national statistics agencies that have ongoing framework agreements with MNOs to access selected CDR-derived data and could facilitate access.

The benefits to such an approach are multifold. Building developing country officials' capacities, skills, and knowledge on mobility data is an important element of paving the future of this work. Obtaining data from all or the largest MNOs in a country can increase the accuracy and representativeness of the study results. Approaching MNOs individually takes significant time, but savings can be generated when a telecom regulator can act as a data aggregator. Finally, working with government agencies can facilitate the necessary clearances for compliance with data regulations and ensure political support for the initiative. This approach was useful where the regulator had already set up agreements with MNOs for sharing data, making it possible to build on the existing interest and collaboration.

Country case study 2. In one country example, data access was facilitated through a government agency that has an existing relationship with MNOs to apply insights from mobility data to a transport sector project. This prior engagement accelerated communications. The first step in the dialogue was to present the COVID-19 use case and raise awareness on the applicability of mobility to health sector analytical products for the country. The government agency readily agreed to the new use case and to jointly generating aggregate data products. In this case, the government agency possessed the infrastructure to receive data from all MNOs in the country. Thus, analytical insights were based on all MNOs' data, which helped to limit statistical biases that can arise when working with solely one operator's data. Where regulators do not have such existing relationships and technical setup for sharing mobility data, formalizing data sharing agreements and eventually accessing the data may require IT systems to be procured and installed by the regulator, slowing any desired rapid response to a sudden-onset emergency.

2.2.3. Working with TPOs

In the context of this work, TPO refers to an organization or institution that has deep technical expertise in mobile phone data usage and can provide related IT system services to MNOs or regulators (e.g., for capturing CDRs). TPOs can be university research departments, for-profit firms as well as non-profit

organizations. The specific TPO with which an interested party can effectively work depends on several factors, such as whether the TPO has existing engagements in the country, how much practical experience the organization has in providing such services and the costs of working with it. Any conflicts of interest with for-profit TPOs would also need to be considered; where a mobility data analytics firm is a TPO under consideration, use of standard procurement rules and contracts at the country or organizational level is recommended.

TPOs often have the responsibility of managing the relationship with the data provider (the MNO or the government agency) and handling the technical aspects of accessing, processing, and analyzing the data. Data access through a TPO is particularly beneficial when the TPO has an existing data access agreement with either the MNO or government agencies, or both.

When this type of collaboration is possible, it leads to large efficiency gains, as indicators already being produced by a TPO can be shared and applied to policy decisions without the MNO or regulator incurring additional costs. An additional key benefit of working through a TPO is the time saved on negotiating data access and deploying the IT systems to process and analyze the data. Nevertheless, in the majority of cases when working with a TPO, it remains necessary to obtain separate permissions from the MNO first.

Country case study 3.   In two country cases, a TPO that had already been engaged with the regulator was able to leverage this existing relationship and facilitate data access relatively quickly. In both countries, the University of Tokyo (UoT) coordinated both the data negotiations and technical work of analyzing the data. In addition to supporting the regulator in carrying out the data processing and analysis, a strong capacity building activity transferred knowledge to the regulator. The TPO's familiarity with the regulator fostered trust and brokered the work in both contexts, significantly reducing transaction costs.

2.3. Data facilitation

Once data access is achieved, the role of a facilitator is to support data analytics. This is a two-step process, starting with the analysis of the raw CDR data and followed by the analysis of the aggregated indicators. This in turn leads to the application of the data to use cases. In this section, we describe the process for data facilitation as well as some important limitations of working with this type of data.

2.3.1. Data Processing and Analysis

Analyzing the raw CDR data is a sensitive endeavor, since even when individual identifiers are removed, a risk of reidentification remains (De Montjoye et al., 2013). De Montjoye et al. (2018) lay out four approaches for responsible analysis of raw CDR data, two of which were incorporated in this work. In one approach, the research team wrote and shared programming code with the technical team at the MNO, which produced the aggregated indicators that were then made available to the research team. In the second approach, the MNO provided remote access to its infrastructure to the research team, which was able to run the analytical programming code on the server provided by the MNO. In both options, the sensitive individual data did not leave the MNO premises, which strengthened data security and protection. A challenge of this approach is that the ensuing analysis is limited to the processing capacity of the MNO as well as on the availability of the MNO's technical team to provide support for setting up the server and managing processing-related errors.

To facilitate these approaches, open source code for a set of CDR-derived indicators was produced by the research team. Since the underlying CDR data are uniform across MNOs and countries, the open source code facilitates and accelerates working with new country cases.[5] It is built on code made available by the TPO Flowminder, which at the beginning of the pandemic produced a set of simple indicators to support MNOs in generating analytics.[6] Additional indicators to support the epidemiological modeling of the disease requested from some government end users were also included.

---

[5] https://github.com/worldbank/covid-mobile-data
[6] https://covid19.flowminder.org/

The ensuing indicators aggregate data across users in space and time, so that the data shared contains no information at the individual level, but only at the geographic administrative level. A limited set of indicators is also produced at the telecom tower level for urban areas. Key indicators are visualized on an interactive dashboard, enabling a geographic lens on population dynamics in the country. The primary indicators offered on the dashboard are measures of density (subscribers per geographic area), mobility (subscribers entering a geographic area, exiting a geographic area, and net movement between areas), and the average total daily distance traveled. All these are displayed at the geographic administrative level. The focus is on change over time from baseline values, which are defined as the average by day of the week across February 1 to March 15. Change is measured as counts, percentage change, and z-score since each of these measures provides a different perspective.[7]

Two risks remain with aggregation: (a) a potential threat to group privacy and (b) the possibility of individual privacy loss even from aggregated data (Pyrgelis et al., 2017). To mitigate these risks, a few strategies were tested. For the aggregated indicators, observations with 15 or fewer SIM cards were removed from the data and marked as missing. This prevents potential reidentification due to small population sizes. In one country case in which a TPO supported the work, differential privacy was used on the aggregated indicators to further limit possible loss of privacy (Dwork, 2008). The aggregated indicators will not be publicly released, instead only the final analyses will be shared. Nevertheless, since the aggregated indicators are shared with relevant government agencies supporting COVID-19 projects, maintaining a high level of privacy and security is critical. Therefore, to ensure safe use of the data, the dashboard showcasing aggregated indicators was reviewed by a technical team for accuracy and by experts from the country for potential sensitivities related to group privacy.

2.3.2. Data limitations

While mobile phone data can provide useful information on mobility that is not available at such a high temporal and spatial resolution from any other data source in developing countries, there are important limitations. There are inherent biases in mobile phone datasets that should be considered when making population-level inferences that may feed into policy making. First, in most low-income countries, mobile phone penetration is not universal, and as such, there is a part of the population that is not represented in the mobile phone data (Silver and Johnson, 2018). There is variability in ownership of phones among different demographic groups based on age, income, and gender as well as on potential geographic differences, all of which affect the representativeness of the data (Frias-Martinez and Virseda, 2012; Wesolowski et al., 2012a). Second, there are some phone usage behaviors that can affect the accurate measure of mobility: in some cases, people use more than one SIM-card with a single device (Goller and Kjetil, 2018), while in other cases a mobile phone is shared by several people (Blumenstock and Eagle, 2010). Additionally, since CDR records only capture behavior when the phone is being used, this could also introduce bias if phone usage is correlated with mobility behavior (Ranjan et al., 2012). Finally, in most cases due to the challenges of accessing mobile phone data, unless it is done for all MNOs centrally through a regulator, often it is only possible to obtain data from one MNO for that country. If the MNO is correlated with specific individual characteristics (such as the income level of the users) or has limited geographic coverage, this could introduce bias.

The limitations are not irreparable as there are ways to alleviate them or at minimum to identify them in order to caveat results. For example, in order to account for the population that does not own phones, adjustment factors can be derived from census or survey data to scale mobile phone data-based estimates to match with the general population (Ricciato et al., 2015). In some countries, differential ownership of phones across demographic groups has not been found to significantly impact analyses of mobility (Wesolowski et al., 2013). Similarly, comparing demographic characteristics of users with different operators in Senegal using survey data, significant differences were

---

[7] For example, a low population area could see a large percentage change in people, but this may not be meaningful from a policy perspective if it only represents a small number of people.
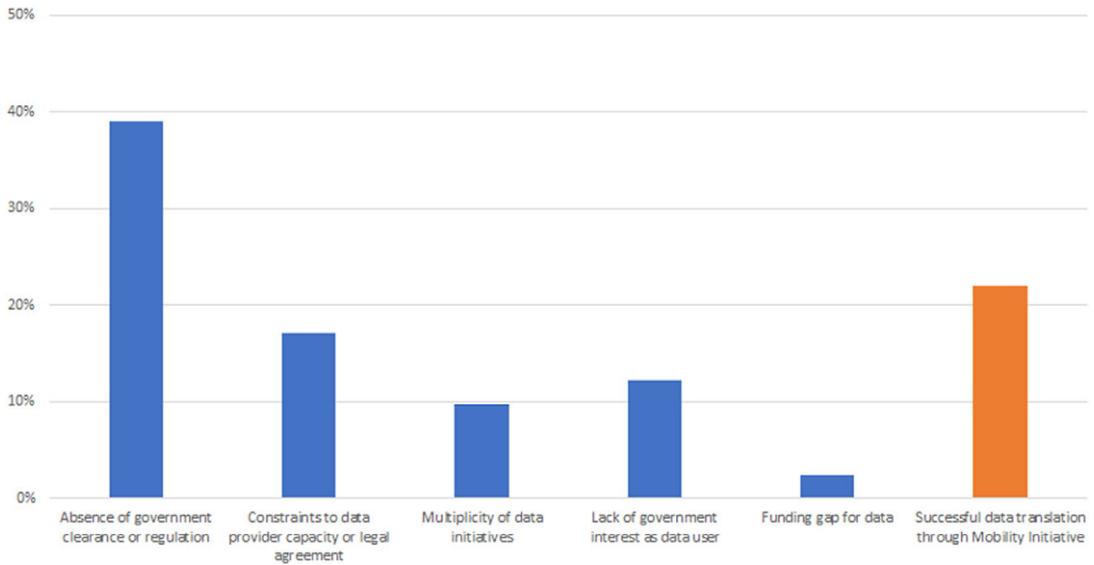
Figure 3. *Main roadblocks to successful data translation and percent of successful cases.*

not found (Milusheva 2020). Nevertheless, using existing survey data on mobile phone ownership and demographic characteristics can be an important addition to mobile phone analyses in order to demonstrate possible biases (Arai et al., 2016). Furthermore, practitioners can collect survey data at a smaller scale to profile usage behaviors across different demographic groups and develop adjustment factors that can be differentially applied, as demonstrated by Arai et al. (2014). Nevertheless, even where all the biases are eliminated, there is an upper limit as to how much mobile phone data can explain/predict the mobility of individuals, and they are not meant to replace conventional survey-based approaches (Song et al., 2010).

3. Outcomes

Out of 41 country cases considered, data access was pursued in 37 cases, and of these, successful collaborations were established in nine country cases. During the process, five main challenges became apparent (Figure 3): (a) the difficulty in obtaining government clearance or reaching regulatory compliance; (b) the necessity to negotiate as of yet non-standardized legal agreements with data providers; (c) a lack of investment and appropriate funding mechanisms; (d) the presence of other facilitators (translators) already working with the government or MNOs — or MNOs already working directly with the government on a limited set of topics; and (e) capacity gaps across government stakeholders. This section describes these challenges, discusses the successful cases and explains their outcomes.

3.1. Data access challenges

3.1.1. Government clearance and regulation

Decision-making between regulatory authorities and the relevant ministries (i.e., the Ministry of Communications) may not be straight-forward in every context, since data protection legislation varies across countries. Part of the labor intensity of securing data access is navigating the sovereign law on a case-by-case basis. Certain legal provisions could impact decisions on engagement, while the absence of a robust data protection framework could impede it. In other cases, exemptions for processing certain types of personal data may be permissible for research or public health purposes and could thus expedite agreements. For example, General Data Protection Regulation (GDPR) makes provisions for the use

of personal data in public health emergencies such as epidemics without permission of the data subject, however certain requirements must still be met regarding anonymization and consent in the context of MNO big data.[8] This may be particularly relevant in contexts where an MNO is headquartered in a country subject to GDPR and may then extend these obligations to its operations in other jurisdictions. The variation of regulatory rules and privacy frameworks across countries can make this onerous to navigate. In almost half of the cases for which data access was attempted, the main challenge was related to government (political) approval based on diverse sets of regulation.[9]

3.1.2. Government interest as a data user

One of the challenges faced was stimulating government ownership. There are a number of possible reasons for this. An important difficulty facing government agencies is competing demands at a time of crisis. Especially during a global crisis such as COVID-19, many potential technological solutions are being proposed to policy makers without clarity on the value added of the different approaches, which can lead them to adopt a limited set of solutions. Another important area is around limited capacity. Given the relatively new opportunity to use these datasets, substantial capacity building is needed for government officials spanning the ministries of communications, telecom regulators, national statistics agencies, and agencies involved in sudden-onset emergencies to utilize the data.

Additionally, as outlined by Abebe et al. (2021), the reluctance to engage in data sharing initiatives may originate from insufficient trust. Data sharing can lead to risks for local communities, especially if data are taken out of context or not analyzed and interpreted with the appropriate local knowledge. Investing time to understand local contexts and build relationships with local research communities that can support and lead the research is important not only for the short-term success of these efforts but also to ensure longer-term sustainability of such initiatives.

3.1.3. Legal agreements with data providers and funding

MNOs own the data to be accessed; data sharing agreements are therefore needed to grant access. In almost 20% of the cases, it was not possible to sign a data sharing agreement with the data providers. Establishing a new data sharing agreement requires trust and time, and developing this between stakeholders during a sudden-onset emergency was challenging or unfeasible. In some cases, establishing a comprehensive legal agreement was possible, but cumbersome. The length of such agreements, due to the number of needed provisions, requires review by lawyers that carries a cost that some providers do not want to incur if they are providing the data pro bono. In other cases, the initial contact was made through a TPO that was already working with the MNO, but it was not possible to leverage this existing partnership to collaborate as the legal agreement between the MNO and TPO was exclusive and prevented the sharing of data with other organizations.

In only one of the cases, the high cost of accessing the data was the main challenge, while in other cases, a legal agreement could not be reached due to the cost implications for the MNO. Funding was therefore an implicit constraint. The substantial costs associated with setting up partnerships as well as collecting, extracting, processing, hosting, and protecting these data require new models of funding. While it is challenging to quantify the level of return for investing time and resources into securing CDR-derived indicators and building these analytical products today, its integration as a key tool for more effective and efficient situational awareness and decision-making should inform models to quantify its value over time. All this points to the need for preparedness — for having data sharing agreements that are agreed upon well before a crisis and can be deployed during sudden-onset emergencies.

---

[8] The GDPR is the primary law regulating personal data protection in the European Union.

[9] In four of the cases, working with this type of data was inadvisable due to security concerns; therefore, after initial discussions, the data were not pursued.

3.1.4. Multiplicity of players

In some cases, though there was an existing collaboration between an MNO and a government agency or a TPO, the MNO was reluctant to launch into a new partnership due to the lengthy process of establishing new agreements and managing simultaneous relationships. While the hope had been that existing collaborations (whether directly with the government or with a TPO) would help to facilitate the new engagements, in some cases, they limited them. For example, in one African country, a TPO had an existing partnership with an MNO supporting the Ministry of Health, yet due to the strictly bilateral nature of the data sharing agreements in place, development organizations could not build on this partnership. This presents a lost opportunity because the same indicators can often serve multiple use cases across different partners. For a number of reasons, though, including the sensitive nature of the underlying data, MNOs prefer to sign bilateral agreements with each institution. Given the lengthy nature of setting up these agreements, it can also mean that once an agreement has been set up with one organization, an MNO may be reluctant to work with others. In one case, several players were working together with the goal of supporting COVID-19 efforts and it was possible for the MNO to sign an agreement with one party and allow the sharing of aggregated data with the other parties for the purpose of the project. Greater investment in and clarity of predefined processes and agreements for different actors could capture some of these missed opportunities (Oliver et al., 2020).

3.2. Ongoing engagements

Within the first seven months of this initiative, MNO approvals to access aggregated mobility data were received for 16 countries out of the initial 41 cases (see Figure 4 for breakdown of country case outcomes). Of these, the necessary government approvals to use the data were obtained for five countries in those first seven months, enabling the production of analytics on COVID-19. An additional four country cases were
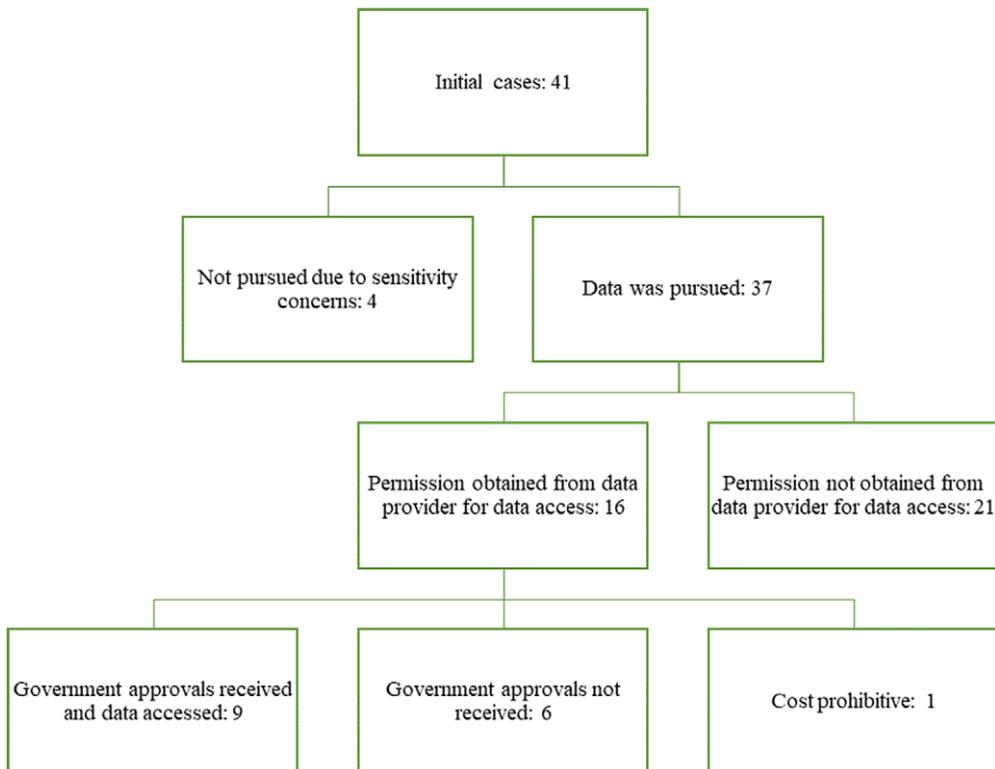

Figure 4. *Diagram of country case outcomes.*

fully approved and realized within 12 months of the initiative start, for a total of nine country cases. Yet, one of the main lessons was that even in a crisis, this type of initiative can take six months to one year to become operational.

In the remaining seven countries for which MNO approval was obtained, the main challenge was obtaining the needed government approvals. In three country cases, presidential elections in 2020 prevented obtaining government buy-in because of the potentially sensitive nature of even aggregated mobility data during election season. In three other cases, discussions are ongoing with the governments. In one case, the cost of accessing the data is prohibitive. With some countries winding down their lockdown measures, the initial epidemiological emergency is no longer as acute. Nevertheless, mobility data use cases for COVID-19 in the medium-term focus on the allocation of resources and vaccines as well as understanding food security needs. With the dramatic surge in cases in India in April and May 2021, it is also clear that epidemiological modeling efforts remain relevant as the pandemic continues worldwide.

Focusing on the nine success cases, Figure 2b shows the main approaches for obtaining data access. For almost half of the cases, accessing the data through an MNO headquarters was effective. There were two MNO HQs that provided access to two countries each where also all the necessary government approvals were received. In two cases, access was provided through the regulator, which aggregated data from multiple MNOs. In two cases, an individual MNO was approached. In terms of speed, working directly with a local MNO was significantly faster for signing an agreement and starting to work. While working with an MNO HQ helped to facilitate more country cases, it required navigating multiple levels of bureaucracy (at both the HQ and local levels), which took significantly longer.

In one case, data were accessed through a TPO, which was able to obtain permission from the counterpart MNO to share the aggregated data. While data access was received directly from a TPO in only one case, collaborations with TPOs were undertaken in five of the other country cases. It remained necessary to coordinate directly with the MNOs or regulator in those cases, however the TPO produced the relevant indicators, thus minimizing duplication of effort and increasing the number of use cases. In two cases, collaboration with other actors, such as the GSM Association (which was working to coordinate similar analyses in these countries), helped to maximize efforts and results from these projects.

For the cases in which all stakeholders aligned—with MNOs willing and able to provide aggregated indicators, all regulations in compliance and an engaged end-user on the government side—several outputs emanated from the alignment. In one country case, the aggregated mobility indicators generated through the mobile phone data were combined with traditional survey data from the census and Demographic Health Surveys to parameterize an agent-based model, which simulates how the virus could spread across and between districts. The model enabled the study of the evolution of case numbers under various policies and scenarios, by modeling the trajectories based on how policies influenced mobility and interactions between people. These models were then shared with the government's COVID-19 Research Group and used along with other modeling and information in considering policies going forward. The model developed could be adapted and applied to the other country settings in which mobility data are available along with other data needed for informing the model.

Additionally, dashboards were produced to demonstrate the change in population movement and density over time and are useful for policy makers to understand the evolution of population dynamics. Figure 5 shows the change over time for one of the mobility indicators that is tracked in the dashboard, demonstrating high variability in movement at the time of the COVID-19 pandemic and the implementation of subsequent policies. In one country case, the data are being integrated into a broader dashboard of indicators that government and donor agencies are using for evaluating the real-time situation. Understanding population dynamics and how they might drastically change during a pandemic as people react to new policies or changes in the crisis can be vital for the government to ensure access to basic needs.

An important element of the work is building capacity within local institutions. To this end, capacity building activities have been conducted and additional activities are planned to train local researchers, PhD students, and government agency officials in the different countries. Technical trainings have been held in
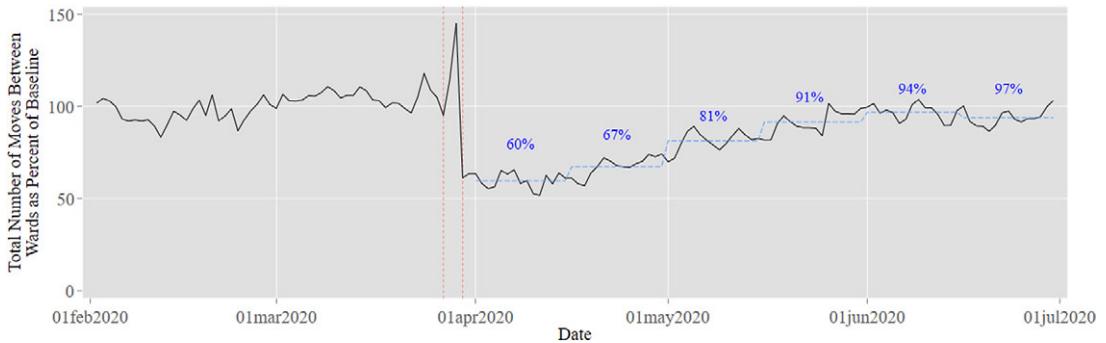
Figure 5. *Percentage change in number of trips between administrative areas in one country example.*

two countries focused on producing indicators from CDR data and visualizing such indicators. These have included discussions of the limitations of these data, which are important to ensure that the analytics are used appropriately and that the risk of wrong conclusions influencing policy is minimized (Zhao et al., 2016; Blumenstock, 2018). Additional trainings are planned on use cases, including epidemiological modeling, with a focus on what type of analytics and modeling are possible with mobile phone data so that policy-makers can advocate for the production of these in the future when the next emergency arises.

In order to facilitate future work in this space, the code for the indicators, the dashboards, the visualizations and the epidemiological modeling is open source on Github, along with training materials that could be applied to new settings.[10] Open source code and resources facilitate the technical aspects of collaborations in the future; this has been championed by TPO organizations such as Flowminder, with many stakeholders standing to gain from code sharing initiatives.[11]

4. Lessons Learned

A history of one-off pilot projects, an absence of standardized practices, a lack of sustainable business models and regulatory barriers have meant that in many cases, aggregated and anonymized mobility data are difficult to access, if they are made available at all. Based on the experience across the 41 country cases, there were seven challenges that required concerted effort to overcome. These encompass the five roadblocks discussed in the previous section and expand to include broader obstacles. They include (a) variation within the ecosystem; (b) insufficient awareness across stakeholders of the value and cost of these data; (c) a lack of investment and appropriate funding mechanisms; (d) a need to build trust between stakeholders; (e) inconsistent approaches to data standards and agreements; (f) a lack of consensus on approaches and models; and (g) capacity gaps across all stakeholder groups ranging from technical capabilities to available human resources to integrate these insights. While each country and MNO engagement presented different pathways to success and roadblocks along the way, the realities below were encountered to some degree in all cases.

4.1. Variation within the ecosystem

There is no one-size fits all approach. The road to success looks different in every country.

- There is variation in the approaches, partnership structure, regulatory environment, organizational structure, incentives, and capacity across different MNOs, countries and use cases. Levels of corporate centralization and decentralization are different among MNOs; depending on the

---

[10] https://worldbank.github.io/COVID19-WBG-Mobility-Task-Force/
[11] https://github.com/Flowminder

organizational structure, different approaches and points of engagement will be required. Common to all is the significant level of effort required to align the stakeholders and the components necessary to achieve a viable implementation. A case-by-case approach to relationship development and securing partnerships and agreements is required, but there are opportunities to gain efficiencies in other areas (such as by establishing standards on minimum data requirements).

- There may be unintentional duplication of efforts between organizations wishing to access and work with the data because of limitations around third-party access and the need for bilateral data sharing agreements between each party.

4.2. Awareness raising among stakeholders

There are several assumptions and knowledge gaps within the development community that need to be overcome to institutionalize and systematize the use of these data. A cultural shift to data-driven policy making is needed.

- Enhancing government understanding of how mobility data can be used for development purposes and for more effective, data-driven policy-making can build political support to institutionalize data use.
- Building an understanding of the complexity of the legal and regulatory environment in priority countries can highlight what needs to be done to move such work forward.
- Raising awareness of the different cost drivers associated with gathering, extracting, processing and hosting CDR-derived indicators can support the operationalization of the work.

4.3. Insufficient investment, valuated benchmark pricing, and appropriate funding

There is a series of market failures that need to be overcome.

- Reputational risk and a motivation to support the communities in which they operate lead many MNOs (despite being for-profit entities) to be motivated to provide free or subsidized access in sudden-onset humanitarian emergencies, while absorbing the costs for doing so. For steady-state access for non-emergency development work, however, a challenge is the significant perception gap between MNOs and the development community as to the value and pricing of mobility data. MNOs view mobility data as their commercial asset with real costs associated with collecting, extracting, processing, hosting, and protecting them, whereas the development community is accustomed to receiving data pro bono, and expects to pay nothing or subsidized/wholesale prices. Short-term funding cycles from the development community do not create the right incentives for MNOs—who are primarily profit-making entities—to move beyond one-off pilot projects, which are - save for a few exceptions - generally of low interest to the industry. While developing a pricing model for public sector and development use cases requires further research, acknowledgment that ongoing access will require investment to cover associated costs is needed, as is aligning longer-term demand for ongoing operational efforts (rather than single bespoke projects) with more stable and predictable supply. The solution probably lies somewhere in the middle — free access during sudden-onset emergencies and a wholesale/subsidized cost structure for steady-state access. The absence of benchmark pricing or consensus on a standard valuation model makes this a challenge to attain; creating a shared value proposition between MNOs and development actors that works for both sides would be a valuable contribution to the effort.[12] Lastly, a viable business model for mobility data analytics as a revenue stream for MNOs from commercial clients would strengthen the availability of such data for the development sector.

---

[12] Guidance on valuation of data for data sharing has started being produced, but additional work is needed in this space (IMDA & PDPC, 2019).

- There is currently no clear path forward to earmarking sufficient investments for this work. This may be a topic of further research in the future and deliberations among the development community and the countries they support.

4.4. Trust

Establishing a new data sharing agreement and a system to generate useful analytical insights requires trust and developing this between stakeholders during an emergency can be challenging. Commercial, humanitarian, and public sector interests may overlap and conflict with individuals' rights, leading to misalignment, or the regulatory environment may lack clarity, contributing to a lack of confidence in the partnership.

- In some countries, the COVID-19 response has led to an unprecedented number of governments requesting the data directly from MNOs. While in some contexts, there may be adequate privacy and transparency around this access, in other places it may be a cause for concern and distrust given the potential for abuse of power. Without streamlined approaches to evaluating and addressing privacy risks, human rights and other ethical considerations, this is unlikely to change in the short-term and could pose a risk to individuals if data are shared without the appropriate protections in place.[13]
- In some cases, government officials may feel circumvented or disempowered if the data are not routed through them. In other scenarios, there may be reluctance on the part of some governments to allow CDR-derived indicators and insights to be shared because of concerns about how they may be interpreted or used by other stakeholders. Overcoming these perceptions and building trust takes time, although one way to facilitate the process is to work closely with local researchers and institutions.
- MNOs who have mature commercial offerings around their CDR-derived indicators may cite pricing structures that create an impression that they are trying to unduly profit from development use cases. While acknowledging that there are costs associated with the collection, extraction and processing of CDR data, further efforts should be placed into collectively defining and differentiating business models that could support access to CDR-derived insights for humanitarian emergencies, as well as work toward alignment on defining steady-state public sector and development use cases.

4.5. Data access and regulation

The pathway to requesting CDR-derived indicators can be complex.

- Identifying the decision makers and right points of contact is important. Decision-making between regulatory authorities and relevant government ministries may not be straight-forward in every context and if requests are not routed correctly, this can contribute to lengthy delays. If agreements need to be built from scratch with a disclosing party in an emergency context, it can be time consuming and hard to advance given competing priorities and associated risks. It is therefore useful for agreements and standards to be prepared and pre-agreed for future sudden-onset emergencies.
- Confidentiality and data protection are paramount for all parties. Undertaking a Privacy Impact Assessment and an evaluation of the relevant privacy frameworks as well as ensuring that the most robust and up-to-date aggregation and anonymization tools are being used are critical steps to building confidence and reducing risk. Because data protection legislation is not uniform, securing data access requires a case-by-case analysis of the country's legal framework. Even in cases where data sharing complied with the legal framework or specific legal provisions were not yet in place,

---

[13] See Jones et al. (2019) for an analysis of ethical use of CDR data in health research, see the Menlo Report (Dittrich and Kenneally, 2012) on a set of ethical principles for research based on digital data.

political clearance was needed to access the data. The stakes may be higher for different stakeholders, for example, MNOs have more to lose (reputational risk, loss of license) if they are perceived to be in violation of regulation.

- Many countries' data protection legislation mandates that the calculations from the CDR data be derived within the sovereign territory of that country, so that CDR data may not leave the boundaries of the country to, for instance, be used in calculations on the Cloud infrastructure of the MNO's global HQ office. These requirements necessitate IT systems and technical infrastructure to be available on the premises of the in-country MNO.
- Robust Cybersecurity legislative and operational frameworks are required to provide the trusted and secure environment in which data analytics can thrive and contribute to research as well as to economic growth. This includes good practice Cybersecurity and Cybercrime legislation, effective Cybersecurity governance at the national and sectoral levels, well-resourced institutions such as Computer Emergency Response Teams, a labor pool with the right digital skills as well as effective operational and technical platforms for prevention, monitoring, and response to threats.

4.6. Lack of standards and consensus

A lack of standardization is a key challenge from the supply and demand side alike.

- Some MNOs prefer to analyze their own data and build analytical products themselves, as they would maintain control of the data itself for privacy, commercial or other reasons. Yet these products may not always be compatible across MNOs in the same country, which prevents the merging of data across MNOs to produce insights that are more representative of the population.
- Some MNOs see an inefficiency in customizing data requests for different partnerships/clients and design off-the-shelf dashboards in the absence of standardized indicators. Yet, different use cases may require different indicator sets; therefore, there may be limits to the usability of ready-made MNO data products. For example, when evaluating the spread of malaria, it is important to know where people have spent the night and for COVID-19, it is important to also know their short-term mobility; therefore, the mobility matrices would be different. The basic indicators for development use cases have not yet been defined or agreed at scale and need to be standardized by consensus across stakeholders.

4.7. Capacity

There is variation across all stakeholders in the level of technical, legal, and analytical competencies required to successfully leverage CDR-derived indicators in both steady-state and sudden-onset emergency contexts.

- In some settings, MNOs may not yet have the technical and analytical capabilities sufficiently developed in-house to produce required CDR-derived insights, or the personnel bandwidth to divert staff time to these initiatives in a timely way. Conversely, in other settings, MNOs may have dedicated units of expertise working with these data and feel best positioned to produce the analytical products themselves.
- Development actors seeking data access and/or CDR-derived insights may not have the capacity or expertise to work with or apply these data and insights to actionable decision-making. An understanding of the broader social, political, and cultural context that surrounds these data is critical to their utility and will require internal capacity building within development organizations as well as investments in partnerships with local organizations, academia, and government.
- In many cases, government agencies' experience in accessing and analyzing big data is still limited. An important role of the translator becomes the transfer of skills and knowledge as well as a vision for the sustainability of the data systems.

5. Ways Forward

Some of the first seminal research on the ability of cell phone data to inform public health response was published in 2012 (Wesolowski et al., 2012b). Over the last 8 years, different actors have completed various successful academic and pilot projects to prove efficacy (Bengtsson et al., 2015; Ihantamalala et al., 2018; Lai et al., 2019; Milusheva, 2020). Yet the systematic integration of these data for development and humanitarian planning and policy at scale has not been realized. The inability of the development community to effectively access mobility data for the Ebola crisis response did not bring us to a point of having resolved the bottlenecks prior to the COVID-19 response. In order to shape a future in which these data can offer the impact they promise for better preparedness and prevention, situational awareness and decision-making, a bold ambition and an institutionally comprehensive approach are needed.

From the analysis of the experiences over one year, there is an emerging model for the future that enables steady-state access that is operationally available over time to inform policy making and development interventions. Laying the foundations for this requires investment in relationship development, awareness raising, stakeholder coordination, demand generation and capacity building, creation and adoption of standards, and a long-term financing plan to fund necessary activities. If these components are invested in and put in place for steady-state access, it will enable a more efficient route to pro-bono access at times of sudden-onset humanitarian emergencies and help build resilience and strengthen response and recovery. Below are considerations for realizing this opportunity:

5.1. Establishing a vision

A bold ambition supported across the development community, development country governments, and MNOs, coupled with long-term investment is required to develop a stronger ecosystem for leveraging mobility data for development outcomes.

- There is a need to think beyond single and smaller-scale pilots and focus on scaling up to big picture opportunities, with associated commitment to long-term planning and resources while considering appropriate measures to mitigate the risks of using CDR-derived indicators. Each country is called to establish an integrated national data system that provides for a sustainable and equitable data sharing ecosystem (WBG, 2021).
- Building capacity within local government agencies and research institutions is critical to ensuring that future efforts for leveraging these data can be led by country-level teams that are best placed to ensure the analyses fit the local context and are integrated into day-to-day policymaking.
- An effort to determine an appropriate and reasonable cost-based pricing structure for mobility data for steady-state development work—which would not be as lucrative as for-profit commercial data services for the private sector—and reasonable standardization of pricing for non-profits would help to standardize a path forward.
- Stakeholders could help to establish the conditions and criteria under which these data can be provided free of charge to produce public good analytical products in sudden-onset emergencies such as pandemics, humanitarian crises, and natural disasters.

5.2. Strengthening the foundations to integrate mobility data into policy and practice

International organizations could use their roles as neutral brokers and conveners to facilitate a global, multi-stakeholder dialogue aimed at establishing or accelerating standardization efforts and defining public sector use cases, as well as enabling trusted research environments. Potential actions include:

- Defining what a "trusted environment" for data sharing might look like on a larger scale, including aspects of data protection (such as secure anonymization and privacy preserving methods), security, and respect of ethical principles.

- Developing a widely-adopted protocol and set of standards and regulatory levers. This could include thought-leadership in establishing consensus on predefined indicator sets for prioritized use cases that could be adopted across the ecosystem and documenting as well as promoting examples of regulatory good practice (such as flexible regulation during sudden-onset emergencies) in enabling access to these indicators for development or humanitarian purposes.
- Creation of standardized and pre-approved templates for licensing/data sharing agreements that could be socialized and adopted as part of steady-state efforts or more efficiently activated for pre-agreed crisis contexts.
- Continued investment in streamlined ethical and legal review and approval processes and guidelines that support consistency, appropriate due-diligence and efficiency. Building on existing efforts such as those by UN Global Pulse (UNDG, 2017) and the Digital Impact Alliance to establish a set of good practice criteria for evaluating any privacy, human rights or associated risk across country contexts would enhance safeguards and help build trust and confidence in the ecosystem.
- Developing frameworks for public–private research partnerships to create collaborative and trusted research environments that allow for the development of algorithms and production of new data products that require integration of data from different data sources.

6. Conclusion

In advanced economies with global MNOs, CDR-derived insights and other data assets have become an important part of an MNO's commercial strategy, and these companies continue to invest in the capabilities needed to harness data for a range of public sector and commercial use cases. Conversely, in many low- and middle-income countries where these data could have the greatest impact, the capacity to process and integrate them is nascent. In some instances, translating the analysis of these data to respond to real world problems and inform decision-making has been elusive, making its value and relevance less obvious. Further compounding the challenge is an absence of cohesive global leadership, coordination, and governance.

The COVID-19 crisis has highlighted both the value of CDR-derived indicators for supporting policy and decision-making, and the challenges in establishing the agreements and obtaining government clearances to secure access. These challenges are surmountable through coordinated efforts, long-term investment, concerted capacity building, and establishment of standards and common approaches. The development community, developing country governments, and MNOs have an opportunity to build a more resilient ecosystem for mobility data for steady-state development, sudden-onset emergencies and humanitarian crises. Learning from the lessons of the Ebola and COVID-19 crises, stakeholders must join to create the processes, guidelines, capacities, and permissive regulatory framework now in preparation for a strong and prompt response to the next crisis in the future. This change is needed.

Abbreviations

CDRs   call detail records
MNO   mobile network operator
TPO   third-party organization

References

Abebe R, **Aruleba K**, **Birhane A**, **Kingsley S**, **Obaido G**, **Remy SL and Sadagopan S** (2021) Narratives and counternarratives on data sharing in Africa. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, pp. 329–341.

Arai A, **Fan Z**, **Matekenya D and Shibasaki R** (2016) Comparative perspective of human behavior patterns to uncover ownership bias among mobile phone users. *ISPRS International Journal of Geo-Information 5*(6), 85. https://doi.org/10.3390/ijgi5060085

Arai A, **Witayangkurn A**, **Kanasugi H**, **Horanont T**, **Shao X and Shibasaki R** (2014) Understanding user attributes from calling behavior: Exploring call detail records through field observations. In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*. New York: Association for Computing Machinery, pp. 95–104.

Bengtsson L, **Gaudart J**, **Lu X**, **Moore S**, **Wetter E**, **Sallah K**, **Rebaudet S and Piarroux R** (2015) Using mobile phone data to predict the spatial spread of cholera. *Science Reports 5*, 8923.

Bengtsson L, **Lu X**, **Thorson A**, **Garfield R and von Schreeb J** (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone data: A post-earthquake geospatial study in Haiti. *PLOS Medicine 8*(8), e1001083. https://doi.org/10.1371/journal.pmed.1001083

Blondel VD, **Decuyper A and Krings G** (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Science 4*(1), 10.

Blumenstock J (2018) Don't forget people in the use of big data for development. *Nature 2018*, 170–172.

Blumenstock J and Eagle N (2010) Mobile divides: Gender, socioeconomic status, and mobile phone use in Rwanda. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. New York: Association for Computing Machinery.

Buckee CO, **Balsari S**, **Chan J**, **Crosas M**, **Dominici F**, **Gasser U**, **Grad YH**, **Grenfell B**, **Halloran ME**, **Kraemer MUG**, **Lipsitch M**, **Metcalf CJE**, **Meyers LA**, **Perkins TA**, **Santillana M**, **Scarpino SV**, **Viboud C**, **Wesolowski A and Schroeder A** (2020) Aggregated mobility data could help fight COVID-19. *Science 368*(6487), 145–146.

De Montjoye YA, **Hidalgo CA**, **Verleysen M and Blondel VD** (2013) Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports 3*, 1376.

De Montjoye YA, **Gambs S**, **Blondel V**, **Canright G**, **de Cordes N**, **Deletaille S**, **Engø-Monsen K**, **Garcia-Herranz M**, **Kendall J**, **Kerry C**, **Krings G**, **Letouzé E**, **Luengo-Oroz M**, **Oliver N**, **Rocher L**, **Rutherford A**, **Smoreda Z**, **Steele J**, **Wetter E**, **Pentland AS and Bengtsson L** (2018) On the privacy-conscientious use of mobile phone data. *Scientific Data 5*, 180286.

Dittrich D and Kenneally E (2012) *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. Washington, DC: U.S. Department of Homeland Security.

Dwork C (2008) Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation.* Berlin, Germany: Springer.

Frankfurter Z, **Silwal A and Seuyong F** (2021) *Measuring Digital Access in Africa. Poverty and Equity Notes*. Washington, DC: World Bank, forthcoming.

Frias-Martinez V and Virseda J (2012) On the relationship between socio-economic factors and cell phone usage. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*. New York: Association for Computing Machinery.

Goller D and Kjetil A (2018) Mobile telephony in emerging markets: The importance of dual-SIM phones. In *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2018: Digitale Wirtschaft*. Kiel, Germany: Leibniz-Informationszentrum Wirtschaft.

GSM Association (2018) Big Data for Social Good: Achievements One Year on and Looking Ahead at Mobile World Congress 2018. Available at https://www.gsma.com/newsroom/blog/big-data-social-good-achievements-one-year-looking-ahead-mobile-world-congress-2018/

GSM Association (2019) Mobile Big Data Solutions for a Better Future.

GSM Association (2020) The Mobile Economy: Sub-Saharan Africa. Available at https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/09/GSMA_MobileEconomy2020_SSA_Eng.pdf

Haddad R, **Kelly T**, **Leinonen T and Saarinen V** (2014) Using Locational Data from Mobile Phones to Enhance the Science of Delivery. World Bank Report Number ACS9644.

Ihantamalala FA, **Herbreteau V**, **Rakotoarimanana FMJ**, **Rakotondramanga JM**, **Cauchemez S**, **Rahoilijaona B**, **Pennober G**, **Buckee CO**, **Rogier C**, **Metcalf CJE and Wesolowski A** (2018) Estimating sources and sinks of malaria parasites in Madagascar. *Nature Communications* 9(1), 3897.

IMDA and PDPC (2019) *Guide to Data Valuation for Data Sharing*. Singapore: Infocomm and Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC).

Jones KH, **Daniels H**, **Heys S and Ford DV** (2019) Toward an ethically founded framework for the use of mobile phone call detail records in health research. *JMIR mHealth and uHealth* 7(3), e11969. https://doi.org/10.2196/11969

Lai S, **Farnham A**, **Ruktanonchai NW and Tatem AJ** (2019) Measuring mobility, disease connectivity and individual risk: A review of using mobile phone data and mHealth for travel medicine. *Journal of Travel Medicine* 26(3), taz019.

Matekenya D, **Espinet Alegre X**, **Arroyo Arroyo F and Gonzalez M** (2021). Using mobile data to understand urban mobility patterns in Freetown, Sierra Leone. In *Policy Research Working Paper, No. 9519*. Washington, DC: World Bank. Available at http://hdl.handle.net/10986/35033

Milusheva S (2020) Managing the spread of disease with mobile phone data. *Journal of Development Economics* 14, 102559. https://doi.org/10.1016/j.jdeveco.2020.102559

Oliver N, **Lepri B**, **Sterly H**, **Lambiotte R**, **Deletaille S**, **De Nadai M**, **Letouzé E**, **Salah AA**, **Benjamins R**, **Cattuto C**, **Colizza V**, **de Cordes N**, **Fraiberger SP**, **Koebe T**, **Lehmann S**, **Murillo J**, **Pentland A**, **Pham PN**, **Pivetta F**, **Saramäki J**, **Scarpino SV**, **Tizzoni M**, **Verhulst S and Vinck P** (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* 6(23), eabc0764.

Peak CM, **Wesolowski A**, **Zu Erbach-Schoenberg E**, **Tatem AJ**, **Wetter E**, **Lu X**, **Power D**, **Weidman-Grunewald E**, **Ramos S**, **Moritz S**, **Buckee CO and Bengtsson L** (2018) Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: Use of mobile phone data. *International Journal of Epidemiology* 47(5), 1562–1570.

Pyrgelis A, **Troncoso C and De Cristofaro E** (2017) What does the crowd say about you? Evaluating aggregation-based location privacy. *Proceedings on Privacy Enhancing Technologies* 4, 156–176.

Ranjan G, **Zang H**, **Zhang Z and Bolot J** (2012) Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review* 16(3), 33–44.

Ricciato F, **Widhalm P**, **Craglia M and Pantisano F** (2015) *Estimating Population Density Distribution from Network-Based Mobile Phone Data*. Luxembourg: Publications Office of the European Union.

Silver L and Johnson C (2018). *Internet Connectivity seen as Having Positive Impact on Life in Sub-Saharan Africa*. Washington, DC: Pew Research Center

Song C, **Qu Z**, **Blumm N and Barabási AL** (2010) Limits of predictability in human mobility. *Science* 327(5968), 1018–1021.

UNDG (2017) *Data Privacy, Ethics, and Protection: Guidance Note on Big Data and the Achievement of the 2030 Agenda*. New York: United Nations Development Group.

Wesolowski A, **Buckee CO**, **Bengtsson L**, **Wetter E**, **Lu X and Tatem AJ** (2014) Commentary: Containing the Ebola outbreak - The potential and challenge of mobile network data. *PLoS Currents* 6(5), 1–11.

Wesolowski A, **Eagle N**, **Noor AM**, **Snow RW and Buckee CO** (2012a) Heterogeneous mobile phone ownership and usage patterns in Kenya. *PloS One* 7(4), e35319.

Wesolowski A, **Eagle N**, **Noor AM**, **Snow RW and Buckee CO** (2013) The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* 10(81), 20120986.

Wesolowski A, **Eagle N**, **Tatem AJ**, **Smith DL**, **Noor AM**, **Snow RW and Buckee CO** (2012b) Quantifying the impact of human mobility on malaria. *Science* 338(6104), 267–270.

Wesolowski A, **Metcalf CJE**, **Eagle N**, **Kombich J**, **Grenfell BT**, **Bjørnstad ON**, **Lessler J**, **Tatem AJ and Buckee CO** (2015a) Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proceedings of the National Academy of Sciences USA* 112(35), 11114–11119.

Wesolowski A, **Qureshi T**, **Boni MF**, **Sundsøy PR**, **Johansson MA**, **Rasheed SB**, **Engø-Monsen K and Buckee CO** (2015b) Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences USA* 112(38), 11887–11892.

World Bank (2021) *World Development Report 2021: Data for Better Lives*. Washington, DC: The World Bank.

Zhao Z Shaw S-L, **Xu Y**, **Lu F**, **Chen J and Yin L** (2016) Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science* 30(9), 1738–1762.