

# Guiding principles to maintain public trust in the use of mobile operator data for policy purposes

Ronald Jansen<sup>1\*</sup> , Karoly Kovacs<sup>1</sup>, Siim Esko<sup>2</sup>, Erki Saluveer<sup>2</sup>, Kaja Sõstra<sup>3</sup>, Linus Bengtsson<sup>4</sup>, Tracey Li<sup>4</sup>, Wole A. Adewole<sup>4</sup>, Jade Nester<sup>5</sup>, Ayumi Arai<sup>6</sup>  and Esperanza Magpantay<sup>7</sup>

<sup>1</sup>United Nations Statistics Division, New York, USA

<sup>2</sup>Positium, Tartu, Estonia

<sup>3</sup>Statistics Estonia, Tallinn, Estonia

<sup>4</sup>Flowminder, Stockholm, Sweden

<sup>5</sup>GSMA, London, United Kingdom

<sup>6</sup>University of Tokyo, Tokyo, Japan

<sup>7</sup>International Telecommunication Union, Geneva, Switzerland

\*Corresponding author. E-mail: [jansen1@un.org](mailto:jansen1@un.org)

The views expressed herein are those of the authors and do not necessarily reflect the views of the United Nations.

**Received:** 24 November 2020; **Revised:** 13 July 2021; **Accepted:** 21 July 2021

**Key words:** COVID-19 response; Fundamental Principles of Official Statistics; mobile operator data; privacy protection; public trust

## Abstract

The COVID-19 pandemic has accelerated the use of mobile operator data to support public policy, although without a universal governance framework for its application. This article describes five principles to guide and assist statistical agencies, mobile network operators and intermediary service providers, who are actively working on projects using mobile operator data to support governments in monitoring the effectiveness of its COVID-19 related interventions. These are principles of necessity and proportionality, of professional independence, of privacy protection, of commitment to quality, and of international comparability. Compliance with each of these principles can help maintain public trust in the handling of these sensitive data and their results, and therefore keep citizen support for government policies. Three projects (in Estonia, Ghana, and the Gambia) were described and reviewed with respect to the compliance and applicability of the five principles. Most attention was placed on privacy protection, somewhat at the expense of the quality of the compiled indicators. The necessity and proportionality in the choice of mobile operator data can be very well justified given the need for timely, frequent and granular indicators. Explicitly addressing the five principles in the preparation of a project should give confidence to the statistical agency and its partners, that enough care has been exercised in the set up and implementation of the project, and should convey trust to public and government in the use mobile operator data for policy purposes.

## Policy Significance Statement

Citizen support for government policies is to a certain degree dependent on public trust in the data used for those policies. During the COVID-19 pandemic, mobile operator data were used to compile timely, frequent and granular indicators for monitoring of the effectiveness of government interventions aimed at reducing human mobility and by extension reducing the spread of the virus. The article presents five principles to guide and assist

statistical agencies, mobile network operators, and intermediary service providers. These are principles of necessity and proportionality, of professional independence, of privacy protection, of commitment to quality, and of international comparability. Compliance with each of these principles can help maintain public trust in the handling of these sensitive data and their results.

## 1. Introduction

In 2020, most governments around the world were faced with the same problem: SARS-CoV-2, a virus, which is spread from person-to-person (WHO, 2020) and could spread rapidly through the population causing a stress on health facilities and causing a potentially large death toll. Government interventions were aimed foremost at stopping the spread by isolating infectious individuals and tracing their contacts, decreasing interpersonal contacts through which the virus could spread, and drastically reducing population mobility to avoid the virus reaching new population groups.

In many countries, national statistical offices<sup>1</sup> (NSOs) were asked by their government to assist in producing indicators, which would inform about the changes of human mobility patterns—as a result of government interventions—and inform epidemiological models about physical distancing and the spatial spread of COVID-19. Traditional surveys (even by Internet or phone) do not give the level of detail or the frequency needed to monitor human mobility. Therefore, NSOs had to look into nontraditional data sources to provide the requested detailed and high-frequency indicators.

Mobile operator data<sup>2</sup> has previously been shown to be highly useful to understand human mobility patterns (González et al., 2008). It has shown very valuable in crisis situations where other relevant data sources are scarce and where there is a strong need for frequently updated information on population mobility (Järv et al., 2012). In infectious disease outbreaks, mobile operator data has shown to improve prediction of the spatial spread of infectious diseases (Bengtsson et al., 2015; Wesolowski et al., 2015), findings which have also shown to be applicable for COVID-19 (Jia et al., 2020). Combined with clinical and public health data, mobile operator data is likely to have an important role in planning rollbacks of government interventions (Kishore et al., 2020). Additionally, mobile operator data has also shown value in certain areas of official statistics, especially for tourism statistics (Saluveer et al., 2020).

The COVID-19 pandemic has accelerated the use of mobile operator data, although without a universal governing framework for its application.

### 1.1. Community of official statistics

At the international level, the work of NSOs is guided by the United Nations Statistical Commission, which sets the international standards for official statistics at its annual meeting for all UN member states, where delegations are led by their Chief Statisticians. In 2014, the Commission created the UN Global Working Group on Big Data for Official Statistics<sup>3</sup> (GWG), consisting of experts from 30 countries and 16 international agencies (United Nations, 2014a). The GWG was requested to give guidance on the use of Big Data for official statistics and develops—among others—methodology as well as practical guidance in the use of various Big Data sources. One of the GWG task teams prepared a Handbook on the use of mobile operator data which provides details on the data sources (active and passive positioning data), access to the data sources, partnership models, and quality assurance (United Nations, 2019a). It

<sup>1</sup> Many countries have a centralized statistical system with a lead role by the NSOs. However, there are also countries, which have a decentralized system, like the United States or India, where the responsibilities over statistical domains have been distributed. For ease of reference, we use the term NSO, which is meant to refer more generally to the whole national statistical system.

<sup>2</sup> In this article, we refer to data produced by mobile network operators as mobile operator data. In academic research mobile positioning data, mobile phone data, CDR/XDR data are also often used and should be considered as subset of mobile operator data.

<sup>3</sup> This intergovernmental body has recently been renamed to United Nations Committee of Experts on Big Data and Data Science for Official Statistics, see <https://unstats.un.org/bigdata/about/index.cshtml>

further develops specialized handbooks on the use of mobile operator data for statistics on tourism, migration, displacement, dynamic populations, commuting, transport, and information society. This task team consists of experts not only from NSOs, but also from private sector, nongovernment organizations and academia. During the COVID-19 pandemic several institutes, which collaborate in this task team, got involved in projects at national level to generate indicators using mobile operator data for the COVID-19 response.

Given the heightened attention to COVID-19 related data and given public awareness that mobile operator data are being used to track human mobility, maintaining public trust in using such data to derive policy indicators is essential. This article describes five guiding principles to assist those actively working on projects using mobile operator data to monitor government interventions. Adhering to these principles should assure the public of the proper use of the mobile operator data. This article is written mostly from the perspective of the NSOs, not only because the team of authors was collaborating under the umbrella of the UN Statistical Commission, but also because of the role society has entrusted to NSOs in using all sorts of data to inform society on relevant emerging topics. Furthermore, the article is focused on data for mobile positioning and does not cover mobile applications for contact tracing.

The five guiding principles have been derived from the Fundamental Principles of Official Statistics<sup>4</sup> (FPOS), which were adopted by the UN Statistical Commission in 1994 to assure trust of the public in the integrity of official statistical systems. Confidence in statistics depends to a large extent on respect for the fundamental values and principles that are the basis of any society seeking to understand itself and respect the rights of its members. In this context professional independence and accountability of statistical agencies are crucial. The importance of the FPOS was recognized by the UN General Assembly with their endorsement at the highest level of the United Nations in 2014 (United Nations, 2014b).

### *1.2. Examples of use of mobile phone data for COVID-19 response*

In March 2020, projects were started at about the same time in Ghana, the Gambia, and Estonia to monitor the effect of government interventions to mitigate and reduce the spread of the virus during the COVID-19 pandemic. In each of those cases mobile network operators (MNOs), the NSO and service providing institutes were involved to compile indicators for the government on the basis of mobile phone records. Such phone records are sensitive and should be treated with the utmost care, since they could reveal information about individuals and their whereabouts. Given the emergency situation the monitoring of human mobility needed to be done quickly, with high frequency and with sufficient geographical detail. The purpose of this article is to provide a checklist of principles, which are recommended to be followed so public and government can maintain trust in the handling of the data and their results even in situations requiring urgent action.

Trust in the treatment of personal data is the cornerstone of the work of NSOs, which have always been entrusted to work with sensitive data, for example, population census data, household surveys, or administrative tax records. The public should have full confidence that the privacy and confidentiality of those data are protected, and that data are only used for statistical purposes (and never for any law enforcement activities, for example). The focus of this article is on the treatment of the mobile operator data in the collaboration between private sector companies, the NSOs and intermediary service providers. Positium,<sup>5</sup> Flowminder,<sup>6</sup> and the University of Tokyo<sup>7</sup> were the intermediate service providers in Estonia, Ghana, and the Gambia, respectively. They provide services in data preparation, development of data

<sup>4</sup> See <https://unstats.un.org/fpos/>

<sup>5</sup> Positium is a data analytics company from Tartu, Estonia, which specializes in mobile positioning data for official statistics, see <https://positium.com/>

<sup>6</sup> Flowminder is a Swedish nonprofit foundation, which provides information and capacity strengthening to governments, mobile network operators, national and international agencies, and researchers in low- and middle-income countries for humanitarian and development purposes, see <https://www.flowminder.org/>

<sup>7</sup> The Center for Spatial Information Science at the University of Tokyo, see <https://www.csis.u-tokyo.ac.jp/en/>

pipelines and the compilation of indicators. Experts of Positium, Flowminder, and the University of Tokyo are coauthors of this article.

## 2. Monitoring Human Mobility to Contain the Spread of COVID-19

In 2019, almost the entire world population (97%) lived within reach of a mobile cellular signal, and the penetration rate of mobile phones was reaching more than 108 per 100 population (ITU, 2019). Because people in general carry their mobile phones with them when moving around, the signals provided by mobile phones give an estimate of how many people are at specific places at a specific time. For this reason, mobile operator data have proven to be helpful to determine the size of daytime and night-time populations, community patterns, or the number of tourists and migrants. In a similar way, these data can be used to check if government interventions led to reduced mobility.

As described by Kishore et al. (2020), data from mobile phones are being used around the world as part of the COVID-19 response, to map population movement, set parameters for disease transmission models, and inform resource allocation. When anonymized and aggregated, these data do not reveal information about individuals and are used to provide epidemiologically relevant estimates about population mobility. For example, the extent to which people are sheltering in place, congregating at parks, grocery stores and transit hubs, and generally moving less (or more) than usual. These data also provide vital insights into travel patterns to help better understand the effect of travel restrictions and the risk of importation from other locations and to inform spatial epidemiological models. These analyses can also be used—while taking precautions not to endanger vulnerable minorities—to identify neighborhoods or communities that could become hotspots for community transmission or that might need additional support to practice physical distancing, or as part of surveillance more generally.

The most common mobile operator data are the call detail records (CDRs), which consists of data entries of active phone use, such as incoming and outgoing calls and sent messages. Less used are data detail records, internet protocol data records, and probed data from signaling information such as location update or cell tower handover. The temporal preciseness of the data improves with every mentioned data type, specifically the more data records there are, the more detailed the information becomes. The spatial preciseness of the data is mainly dependent on the distribution of mobile network cells, which in turn are determined by the population density and pattern (Ahas et al., 2007). Hence, the data are geographically more accurate in densely populated urban areas and near major roads but less accurate in rural areas.

Flowminder<sup>8</sup> published on its website how to use CDR aggregates to derive mobility indicators, such as Crowdedness (how busy and how often busy places are), Population mixing (how many different people visit the same place over time), and Intraregional travel (quantity of movement within a region). Full details are provided on the variables and methods used. This shows at a very practical level how mobile phone data are used to derive indicators on human mobility.

## 3. Principles to Maintain Trust in Official Indicators

How can you ensure and maintain trust in the proper handling of mobile phone data for the compilation of indicators, which are used to inform public policy matters? What would you need to do to earn that trust? This article proposes five principles as a checklist of what you need to do. Those principles have been chosen in relation to the FPOS, which have proven their value for more than 25 years. These are principles of necessity and proportionality, of professional independence, of privacy protection, of commitment to quality, and of international comparability.

There is evidence of a positive relationship between trust in institutions (like the NSO) and citizen support for government policy. Understanding institutional trust is essential to understanding government effectiveness and the functioning of democratic systems of government (OECD, 2017). Questions which

<sup>8</sup> See <https://covid19.flowminder.org/home>

can be posed in relation to the careful handling of sensitive data are (a) Is it really necessary to use mobile phone records? (b) Are those handling these data independent and accountable? (c) Is privacy of the records protected? or (d) Are the results obtained trustworthy and of good quality? Of these questions, the one on privacy protection is generally getting most attention by all parties, as it stands in the limelight of public debates.

The considerations regarding adhering to the principles will be about striking the right balance between fit-for-purpose, privacy, independence and quality. For example, fully anonymized and aggregated mobile phone data protect individual privacy very well but may not allow for performing all the desired analyses required for monitoring human mobility (data may not be fully fit-for-purpose). If the MNO would do all the processing and provide the aggregated data to the other partners in the project, then this would ensure privacy but may raise questions on how the MNO performed the processing even if instructed by the other partners in the project (principles of independence and quality). In all three country cases, which will be described in [Section 4](#), the working arrangement was in fact just that: guided by instructions of the project partners the MNOs pre-processed, anonymized, and aggregated the basic mobile operator data, which were subsequently compiled into final indicators by the NSO and the other project partners.

The FPOS touch on all the aspects of fit-for-purpose, privacy, independence and quality, they underpin the work of the statistical community and they are therefore the main reference for this article. The FPOS state—among others—that official statistics are an indispensable element in the information system of a democratic society, that statistical agencies need to decide independently and only according to professional considerations to retain trust in official statistics, and that individual data are to be treated strictly confidential and used exclusively for statistical purposes. Since the FPOS were recognized at the highest political level, governments should support the national statistical agencies in their efforts to implement the FPOS and provide trusted official statistics as reliable, objective and independent information.

Outside of the statistical community, data protection and privacy are of great concern for the society, and governments have therefore established laws and regulations to protect the rights of its citizens in that respect. At European level a major effort was made in the development and adoption of the General Data Protection Regulation<sup>9</sup> (GDPR), which has consequences within and beyond Europe for the way in which organizations can deal with personal and other sensitive data. Many national data protection authorities (not only those within the European Union) take the GDPR as guidance in the decision-making process regarding the use of personal data. Furthermore, the private sector drafted guidance on the protection of privacy. For example, the GSMA published COVID-19 Privacy Guidelines for MNOs when considering requests for access to mobile operator data in response to the spread of COVID-19 (GSMA, 2020).

This article recommends that a project, which uses mobile operator data to derive policy indicators, should try to comply with the following five principles.

### ***3.1. Principles of necessity and proportionality (fit-for-purpose)***

Official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honor citizens' entitlement to public information [FPOS principle 1]. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs, and the burden on respondents [FPOS principle 5].

The principle here is that in order to justify the use of mobile operator data (a) the derived indicators need to be relevant for the society, and (b) the data source is chosen on considerations of quality, timeliness, costs, and the burden on respondents.

<sup>9</sup> The General Data Protection Regulation is contained in Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. See [https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en)

The relevance of indicators for society could be up for public debate in many instances. However, in an emergency situation, when lives are at stake, the need and relevance of indicators become obvious quickly. In the absence of a vaccine, the only way to stem the spread of the virus was to reduce mobility and contact among the population. Monitoring human mobility requires frequent and geographically granular indicators. Such indicators cannot be accurately obtained with regular surveys or from administrative sources. Mobile operator data, on the other hand, can give timely, frequent, and detailed indicators. These data are generated automatically and therefore do not pose a burden on respondents. The MNOs will carry some burden in making data available and may incur costs in doing so. However, in view of the emergency and the inadequacy of standard data collection tools, mobile operator data are necessary and proportional, and therefore fit-for-purpose. Surveys could still be used as an additional data source to benchmark the results and to get more qualitative information.

### **3.2. Principle of professional independence**

To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage, and presentation of statistical data [FPOS principle 2].

This principle states that statistical agencies should develop, produce, and disseminate statistics or indicators without any political or other interference or pressure from other government agencies or policy, regulatory or administrative departments and bodies, the private sector or any other persons or entities. Such professional independence and freedom from inappropriate influence ensures the credibility of official statistics. This should apply to the NSO as well as to other producers of indicators for official purposes. Statistical releases should be clearly distinguished from political or policy statements. Credibility and independence include also reporting of unexpected results or failure, which can become a reality in every stage of a project with mobile phone data—during data access, input data validation, processing, or output validation. Unexpected results or failure should be documented and shared to create the basis for learning, internally and externally.

Professional independence is crucial for the credibility of the NSO, which is why the NSO needs to take a leading role in the project to ensure that the indicators are developed, produced, and disseminated without any political or other interference. The working arrangements between the NSO and the other actors in the project should further include assurances that scientific principles and professional ethics are strictly followed.

### **3.3. Principles of privacy protection**

Since the actors in the projects under consideration are usually from various stakeholder communities (the NSO, nongovernment organizations, academia, and private sector), we approach the principles of privacy protection from three perspectives:

- (i) privacy protection by the NSO,
- (ii) privacy protection according to the national data protection authority, and
- (iii) privacy protection by private sector entities.

Fundamentally, privacy is protected as a human right. Article 12 of the Universal Declaration of Human Rights<sup>10</sup> states “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks.” Against that background, the national data protection authority aims to put safeguards in place against unwanted use of personal or otherwise sensitive data. The before-mentioned GDPR can serve as guidance for such safeguards. Specific directives have further been

<sup>10</sup> See <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

developed for specific sectors; for example, Directive 2002/58/EC<sup>11</sup> of the European Union (known as the ePrivacy Directive) has been adopted for the electronic communication sector to protect the right to privacy with respect to the handling of personal data in this sector.

Even if the country projects described in this article follow the strict rules of privacy protection of the Statistics Law, they are still subject to national data protection regulations. For example, in Europe processing mobile operator data falls under the ePrivacy Directive. While the GDPR permits processing personal data for statistical purposes,<sup>12</sup> the ePrivacy Directive only permits the processing of electronic communications metadata for limited purposes,<sup>13</sup> and there is no exemption for statistical processing.

Further, these national data protection regulations may be different depending how the mobile operator data have been pre-processed. This is, for example, the case for pseudonymized and anonymized data. Pseudonymized data means that data remain personal, but that direct personal identifiers (e.g., name, ID-code, address, or mobile number) have been replaced with codes that cannot be linked back to a person without additional information and substantial effort. Anonymization goes one step further than pseudonymization and uses disclosure control techniques (e.g., micro aggregation and local suppression) such that it is not possible to identify persons from the data directly or indirectly.

### 3.3.1. Principle of statistical confidentiality and data security

Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes [FPOS principle 6].

The two parts of this principle are equally important. Data are to be kept strictly confidential, which means that the data cannot be made available to anyone else than those who are working at the statistical agency and who have sworn the oath of secrecy. Moreover, the use of the collected data is exclusively for statistical purposes. This means that the data cannot be used for any other purpose, for instance the data cannot be used for administrative purposes or for law enforcement purpose. Population Census data, for example, should never be used to enforce any legal action.

When considering the type of data necessary for statistical analysis, statistical agencies should always consider whether their goals can be achieved using aggregated nonidentifiable data instead of identifiable data. The use of identifiable data, including pseudonymized data, increases privacy risks and requires additional scrutiny against relevant data protection laws, as well as any sector-specific regulation. The principle of necessity and proportionality crosses here with the principle of protection of privacy: if aggregated nonidentifiable data are sufficient to obtain valid indicators for the purpose a project, then abstain from using more detailed identifiable data.

Furthermore, physical, technological, and organizational provisions should be in place to protect the security and integrity of stored data as well. In the case of mobile operator data, it is in general preferred if the MNOs retain full control of their data and process the data within their own internal secure systems to minimize the possibility of breaches in data security.

### 3.3.2. Data protection regulations

There are over 130 data privacy laws in place around the world (UNCTAD, 2020). Most of these laws apply to the collection and processing of personal data and are based on the same commonly accepted privacy principles. These privacy principles took shape in the early 1980s, notably through the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (OECD 2013) and the Council of Europe Convention 108 on the protection of individuals with regard to the processing of personal data.<sup>14</sup> In the European Union, these principles are now enshrined in the GDPR, which is widely seen as a go-to guide to privacy worldwide.

<sup>11</sup> See <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002L0058&from=EN>

<sup>12</sup> See Recitals 156, 162, and 163 and Articles 5(1)(b), 5(1)(e), and 89.

<sup>13</sup> See Article 6 of the ePrivacy Directive.

<sup>14</sup> Convention 108 and Protocols: <https://www.coe.int/en/web/data-protection/convention108-and-protocol>

The GDPR permits further processing of personal data for historical, statistical, or scientific purposes provided that appropriate technical and organizational data safeguarding measures are implemented. If data is anonymized and aggregated, care should be taken to ensure that this data cannot be re-identified through any reasonable effort.

According to the European Data Protection Board (EDPB, 2020), evaluating the robustness of anonymization relies on three criteria:

- (i) singling-out (isolating an individual in a larger group based on the data);
- (ii) linkability (linking together two records concerning the same individual); and
- (iii) inference (deducing, with significant probability, unknown information about an individual).

In the EU, mobile operators must also comply with the ePrivacy Directive (the before-mentioned Directive 2002/58/EC), which includes sector-specific privacy and confidentiality requirements for providers of electronic communications services, including telecommunications providers. Both the GDPR and the ePrivacy Directive exempt data that is anonymized and aggregated. This is true in most jurisdictions with data protection laws: if data is truly anonymous and cannot technically be linked to any individual, the use of that data will fall outside the scope of a data protection law.<sup>15</sup> Pseudonymized data, on the other hand, could potentially be linked back to a person with additional information and substantial effort. In individual cases the national authorities need to decide, if enough safeguards are in place to be confident that re-identification is unlikely.

While the EU has developed a set of rules and specific, timely regulatory guidance, this is not always the case in other jurisdictions. It is important to recognize that some countries are still developing a relevant privacy law or telecommunications law with data processing requirements. The legal landscape may be unclear. In such a situation principles and guidelines provide helpful parameters to facilitate leveraging mobile data while also preserving privacy.

### 3.3.3. GSMA privacy principles

The GSMA COVID-19 Privacy Guidelines (GSMA, 2020) reflect existing mobile industry privacy practices such as the GSMA Mobile Privacy Principles,<sup>16</sup> and provide recommendations on how the mobile industry may maintain trust while responding to governments and public health agencies that have sought assistance in the fight against COVID-19 through the use of MNO data. It is a collection of approaches that the GSMA recommends for MNOs when considering requests for access to MNO data in response to the spread of COVID-19. In practice, such requests may include both personal data, including communications metadata,<sup>17</sup> and aggregated nonidentifiable data.<sup>18</sup>

The GSMA privacy guidelines (GSMA, 2020) say that while, in all cases, data should not be shared except as consistent with law, MNOs should, as appropriate, seek assurances from the Governments or Agencies that they will:

- Use data in a lawful and fair manner.

<sup>15</sup>Note that the use of mobile network operators' data may fall under a number of other sectoral requirements, including telecommunications privacy and security regulations, confidentiality/wiretapping laws, and license conditions.

<sup>16</sup>The GSMA Mobile Privacy Principles are a set of principles intended to guide the mobile industry; the principles are generally aligned to the commonly accepted privacy principles.

<sup>17</sup>As defined in the GSMA COVID-19 Privacy Guidelines: "Metadata" means traffic data including call detail records from mobile networks including where key identifiers such as the mobile number and subscriber information have been replaced with a pseudonym.

<sup>18</sup>As defined in the GSMA COVID-19 Privacy Guidelines: "Aggregated Non-Identifiable Data" means Metadata in an aggregated form with appropriate thresholds (e.g., regarding the number of individuals, time, and/or space) designed to prevent the possibility of individuals being re-identified. This typically includes origin–destination matrices or footfall information generated from Metadata. Although every effort is made to avoid the possibility of re-identification in the way these datasets are designed, there remains a residual, theoretical risk making it difficult to pass the legal test of true anonymity imposed in some jurisdictions.

- Be as transparent as possible with the public about the use of mobile operator data and the applicable legal framework.
- Clearly describe the purpose for which the original data is being shared and follow that purpose throughout the project.
- Conduct a data privacy impact assessment (DPIA).<sup>19</sup>
- Enforce a strict rule prohibiting the re-identification by staff or researchers.
- Have appropriate technological and organizational measures in place to secure all mobile operator data when “at rest” as well as “in transit.”
- Respect principles of equal protection under law, and not to use mobile operator data or analytical insights to discriminate improperly against individuals or groups, or to violate fundamental rights.
- Vet researchers and the scope of their research projects appropriately and bind them to an equivalent standard of adherence to these guidelines.
- Delete any individual level mobile operator data after a defined period or once it is no longer needed for the agreed purpose.
- Be accountable and provide evidence that they have acted in accordance with the assurances given. Establish an independent oversight board to monitor adherence to these principles.

Compliance with applicable law refers both to protecting data privacy but also to making data available to the statistical authority to inform public policies. An agreement between the MNO and the NSO needs to find the right balance. Accountability (mentioned as last point of the GSMA privacy guidelines) is particularly relevant in any data sharing context. Accountability means that parties should be able to demonstrate that they have acted responsibly and complied with relevant law and have ensured that any third parties with whom they have shared data are bound by the same standards. Data should be treated consistently across the chain of custody. Given legal limitations and the need for accountability, the MNO may be hesitant to agree on sharing data, unless certain conditions apply, whereas the NSO will be asking for utmost transparency in the data preparation and sharing of information. The negotiations on the agreement of use of data between the MNO and the NSO in Ghana, as well as in the Gambia took several years to be successfully completed.

### 3.4. Principles of commitment to quality

To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics [FPOS principle 3].

In other words, statistical agencies should be dedicated to assuring quality in their work, and systematically and regularly identify strengths and weaknesses to continuously improve process and product quality. Indicators on statistical output quality are regularly measured, monitored, published, and followed up to improve statistical products and processes.

A key concern with many Big Data sources, such as mobile operator data, is the *selectivity* or the *representativeness* of the dataset. As explained by Buelens et al. (2014): “A subset of a finite population is said to be representative of that population with respect to some variable, if the distribution of that variable within the subset is the same as in the population. A subset that is not representative is referred to as selective.” In other words, units included in the mobile operator data could differ from units missing from the data set. To assess the difference, it is useful to consider the “profile” of the included and excluded

<sup>19</sup> A DPIA is mandatory in Europe for projects where people’s location data is used and otherwise highly recommended to assess the Project in relation to the purpose it is intended for and the rights of data subjects, even when data is pseudonymized: for background [https://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=47711](https://ec.europa.eu/newsroom/just/document.cfm?doc_id=47711), for an example [https://ec.europa.eu/energy/topics/markets-and-consumers/smart-grids-and-meters/smart-grids-task-force/data-protection-impact-assessment-smart-grid-and-smart-metering-environment\\_en](https://ec.europa.eu/energy/topics/markets-and-consumers/smart-grids-and-meters/smart-grids-task-force/data-protection-impact-assessment-smart-grid-and-smart-metering-environment_en) and for an overview see <https://gdpr.eu/data-protection-impact-assessment-template/>

units. For example, what are the differences between those who own a mobile device and those who do not (by geographic location, age, income, etc.)? Selectivity can become a larger problem if only one MNO is included in the project. The *accuracy* of statistical information is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. The *validity* of a dataset is the extent to which it measures what the user is attempting to measure. For example, a survey may directly ask where someone lives, whereas an estimation of the usual place of residence can only be inferred from mobile operator data based on patterns of data points, such as the data points associated with night-time location and other considerations. For projects using mobile phone data to estimate policy relevant indicators, sources, methods and procedures should be described according to scientific standards, so that they are *verifiable* by various stakeholders. Relevant metadata should be made available about the source, processing and results. Typical examples of *reproducibility* comprise data, code and text files, often organized around a source document. Flowminder showed that this can be done by making its source code available, which was used for the calculation of COVID-19 indicators based on CDRs.<sup>20</sup> The benefit for service providers to open up the source code is that others can help in testing and further improve the algorithms.

### 3.5. Principle of international comparability

The use of international concepts, classifications and methods by statistical agencies in each country promotes the consistency and efficiency of statistical systems at all official levels [FPOS principle 9]. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries [FPOS principle 10].

The authors of this articles are all members of the UN Global Working Group on Big Data for Official Statistics,<sup>21</sup> which is developing—among others—internationally agreed practices in the use of mobile phone data for various domains of official statistics. Sharing methods and practices through this kind of multilateral cooperation could help develop and test over time internationally agreed indicators of human mobility.

## 4. Use Cases of COVID-19 Monitoring

### 4.1. Estonia

#### 4.1.1. Project setup

Mobile operator data has been used for official statistics in Estonia since 2008, for tourism and travel statistics (Saluveer et al., 2020) in collaboration between Positium and the Central Bank of Estonia. When—on March 12, 2020—the government declared an emergency situation in Estonia caused by COVID-19, several restrictions on population movement were imposed and the public debated whether there must be a data-driven decision-making process, relying on mobile operator data.

On March 17, 2020, the Director General of Statistics Estonia started discussions on how to derive mobility information from the mobile operators' data to better manage the emergency situation. The government's crisis committee approved the initiative and requested Statistics Estonia to work with MNOs to capture near-real-time data and calculate mobility statistics with daily reporting to the government. Statistics Estonia and Positium developed the methodology applying the activity space method of Positium. Further, the terms of reference were negotiated between the MNOs (Telia, Elisa, and Tele2), the Estonian Association of Information Technology and Telecommunications, the Government Crisis Committee, the Data Protection Inspectorate, the Ministry of Economic Affairs and Communications, and the Chancellor of Justice.

<sup>20</sup> See <https://github.com/Flowminder/COVID-19> and <https://covid19.flowminder.org/cdr-aggregates>

<sup>21</sup> As mentioned earlier, this intergovernmental body has recently been renamed to United Nations Committee of Experts on Big Data and Data Science for Official Statistics, see <https://unstats.un.org/bigdata/about/index.cshtml>

#### 4.1.2. Objectives

The public requested a data-driven decision-making process for implementing and maintaining government restrictions, and the best technology to monitor the effectiveness of movement restrictions was to use mobile operator data. The Crisis Committee asked Statistics Estonia to monitor whether and to what extent the imposed measures of movement restrictions during the emergency were actually observed.

The following mobility indicators were calculated:

- a) Maximum distance from main location (usual residence) over all trips.
- b) Average distance from main location.
- c) The amount of distance traveled as the sum of maximum distances from the main location for each individual trip.
- d) Standard deviation of distances from main location.
- e) Percentage of time spent at the main location.
- f) Number of trips per day.

The Crisis Committee of the Republic of Estonia was to receive a daily data update and a weekly overview based on their request.

#### 4.1.3. Implementation

During the last days of March, Statistics Estonia and Positium compiled a methodology in cooperation with all three MNOs to assess people's daily movements. While GDPR allows analysis of pseudonymous and nonaggregated data for statistics, the Electronic Communications Act (ECA) only allows the processing of anonymous and aggregated location data. Since there was not enough time to work out an agreement with the MNOs on using nonaggregated data, it was decided to retain only the mobility monitoring mission requesting the MNOs just to carry out a simplified calculation to be aggregated and forwarded to Statistics Estonia. Statistics Estonia's role as an intermediary (between MNO and the project team) was crucial as the MNOs did not want to show each other the regional distribution of their customer base.

The results were generalized at the level of local municipalities and, when possible, at a more detailed level (e.g., by urban regions in Tallinn and Tartu). Statistics Estonia reviewed the aggregate data received from the mobile operators and calculated the rate for staying local.

The first data transmission took place on April 1. Comparing the data of the operators revealed problems in the calculation of some indicators, about which Statistics Estonia provided feedback and the operators quickly corrected their algorithms and sent new tables. As a result of good cooperation, Statistics Estonia published the first summary on the mobility of persons already on April 3 (Statistics Estonia, 2020a). A more detailed overview was published on April 9 (Statistics Estonia, 2020b), that is, 4 weeks after the start of the emergency. On the same date, the data was also published in the map application prepared by Positium,<sup>22</sup> which was periodically updated.

#### 4.1.4. Adherence to the principles

- *Principle of necessity and proportionality*: Due to urgent needs to monitor people's movement during an emergency situation the mobile operator data was the best and almost only available data source. It was necessary to have information as close as possible to real time. Compared to weekly household surveys the mobile phone data could give daily information with a 2-day time lag that covered most of the population over 7 years old.
- *Principle of professional independence*: As stated for Statistics Estonia by EC Regulation 223/2009 on European statistics and Official Statistics Act, professional independence means statistics must

<sup>22</sup> An application of Positium for the assessment of mobility commissioned by the Crisis Committee, see <https://liikumisanalyyks.stat.ee/>

be developed, produced and disseminated in an independent manner, particularly as regards the selection of techniques, definitions, methodologies and sources to be used. Statistics Estonia strictly follows the principle. In this project Statistics Estonia was working directly with the three MNOs and was responsible to integrate the MNO data while not to disclosing any of the MNO's business secrets.

- *Principles of commitment to quality* were followed by input and output data quality assurance:
  - Positium provided a list for the operators to check the input data daily. It should be noted, however, that the process of validating input data was out of Statistics Estonia's control who received no information about input data quality and this could be improved with future projects.
  - The output data quality control consisted mainly of comparing the number of mobile phones with the previous days and checking the presence of all the areas. A comparison of the results of the three operators revealed possible errors, in which case the operators were again asked for data. The initial manual quality control became automated at the end.
  - Automatic Controls checked for logical consistency, for example, examined file dates, the completeness of tables, duplicates, total and averages of subscribers by operator per area. Only when everything matched were the output tables automatically forwarded.
- *Principles of privacy protection* were guaranteed because the aggregation of basic phone data was done by the MNOs. The aggregated data, which were received from the network operators by Statistics Estonia for further processing, was completely anonymous and was protected according to physical, technological, and organizational principles used by Statistics Estonia. Data was deleted 1 week after the finishing of the project.

#### 4.1.5. Results

The Crisis Committee received the first overview 14 days after the request was sent. Based on mobile data, the findings after the state of emergency were as follows:

1. Significant increase in the population presence within the surroundings of the main cell tower;
2. Significant decrease in distance traveled from the main cell tower;
3. Significant reduction in the time spent outside the area of the cell tower.

The mobility analysis revealed that compared to the first weeks of the emergency situation when movement decreased, the additional restrictions have not had a clear impact on mobility.

Based on these points, the Crisis Committee also noted the effectiveness of the movement restrictions and, monitoring the number of COVID-19 infections in addition to the movement data, did not consider it appropriate to apply additional restrictions (except for closing individual hotspots) and prepare for easing and ending the emergency. It must be said that the decision was not made just on the basis of a single source. The project was terminated on May 17 with the end of the emergency situation.

#### 4.1.6. Summary

The project was a success in terms of the short time that stakeholders managed to agree on a methodology, started calculating the necessary indicators on a daily basis for all mobile operators, set up data capture and processing, and launched a map application. The result was achieved because of good cooperation with the MNOs. Efficient communication resolved many detected problems quickly. All parties worked to help the country with the necessary data in a difficult situation.

On the other hand, swift action also caused problems. Obtaining aggregated data for Statistics Estonia was the only option, but it significantly reduced the reliability and interpretability of the data. The methodology was general and it could happen that MNOs and Statistics Estonia had a different interpretation of the calculation of indicators. Some errors were detected by checking the magnitudes of the averages and comparing the indicators of different mobile operators. However, it is not possible to detect small deviations from the aggregated data, as the indicators of mobile operators may differ due to

differences in customer bases. It is important to write out the methodology and calculation algorithms in as much detail as possible, using calculation formulas and sample datasets. If possible, the calculations should be tested with common test data and extensive data quality reports should be made.

This was a temporary solution and therefore there were occasional problems with the timely transmission of data. The timing of data transmission by operators varied, making it difficult to automate data acquisition and processing. Should a longer-term cooperation with a similar methodology be required, it is necessary to automate the sending and receiving of data at all parties.

*4.1.7. Data availability statement.* In line with international comparability principle the developed methodology of mobility indicators is not country specific and could be used in other countries or results compared with other mobility analysis. The aggregated mobility data and general methodology are described on Statistics Estonia's website.<sup>23</sup> Any other data that was used within the project has been deleted by involved parties as it was agreed at the beginning of the project.

## 4.2. Ghana

### 4.2.1. Project setup<sup>24</sup>

A formal partnership between Ghana Statistical Service (GSS), Vodafone Ghana and Flowminder has been in place since the start of 2019. In this arrangement Flowminder is working with GSS to produce mobility indicators from aggregated and anonymized data provided by Vodafone Ghana. The partnership was set up with the aim of building capacity within Ghana to enable aggregated, anonymized Call Detail Records (CDRs) to be incorporated into the production of official statistics. Vodafone Ghana prepared with support of Flowminder the aggregates by processing the CDRs of many individual subscribers into an output that characterizes the behavior of the entire group of subscribers. The project was funded by the Vodafone Foundation and the William and Flora Hewlett Foundation.

Prior to the onset of the COVID-19 pandemic at the start of 2020, the project had already progressed significantly, with the partners being in a position to start identifying use cases and producing outputs. In particular, the partners had developed the necessary legal and partnership frameworks to enable the project to proceed. This included engaging with the Ghanaian Data Protection Commission to ensure compliance with local data protection legislations, as well as ensuring compliance with the EU GDPR legislation. The technical infrastructure enabled Flowminder staff to analyze the data, via a secure remote connection, so that all data was processed, physically, within Vodafone Ghana's premises. The technical and governance frameworks ensured that the principles of confidentiality and data security would be strictly adhered to. Only anonymized, aggregated CDR data, that had been approved by Vodafone Ghana, would be made available to the government and other data consumers, with no information about individual data subjects being accessible.

Dissemination activities, including a media event, had been conducted before the epidemic in order to ensure that there was clear, transparent communication about how the mobile operator data was being used in the partnership, and the purpose for which the data were being used. The COVID-19 pandemic presented an opportunity to leverage the infrastructure of the existing project, to derive information about population movements in Ghana from the CDR data used in the project.

### 4.2.2. The objectives

Movement restrictions, closure of schools, banning of public gathering, lockdown of major urban centers were proclaimed in Greater Accra Metropolitan Area and the Greater Kumansi Metropolitan Area in

<sup>23</sup> Firstly, mobile operators identify the main area of each anonymous mobile number, which is used to locate the signal of the mobile masts. The accuracy of the area depends on the position of the mast and is 500 m (in the city) to 3000 m (in the country). The position shall take into account the daily average, maximum, and standard deviation. Secondly, the time during which the anonymous mobile number is away from its primary location is determined. See <https://liikumisanalyys.stat.ee/>

<sup>24</sup> For further details of the project, see Li et al. (2021).

Ghana on March 30, 2020. GSS sought to provide the public and the government of Ghana with information on the effects of these restrictions on social gatherings and internal mobility patterns.

#### 4.2.3. *The implementation*

Given the existing collaboration and infrastructure that was in place, analysis work could start rapidly. Just 4 days after the government's proclamation of mobility restrictions the partners published the first report with insights on changes in mobility based on analysis of the CDR data (Ghana Statistical Services, Vodafone Ghana and Flowminder Foundation, 2020a). The first report included analyses on subscriber presence within regions (administrative level 1) and districts (administrative level 2), and travel between regions and between districts. The analyses covered the time period starting 4 weeks before initial social-distancing and mobility restrictions were introduced, until a March 31, 2020 after lockdown measures had been introduced. This showed how population movements changed in response to each set of restrictions. A second report was released 3 weeks after lockdown measures had been lifted (Ghana Statistical Services, Vodafone Ghana and Flowminder Foundation, 2020b). This contained similar analyses to the first report, but extended results to the post-lockdown period. The report is available from GSS's website<sup>25</sup> and both reports were shared with the Ministry of Health and the Office of the President.

#### 4.2.4. *Adherence to the principles*

In developing the mobility analyses contained in the official reports, the principles above were addressed as follows:

- *Principle of necessity and proportionality*: Owing to the start of the project since 2019, the relevant data pipelines had already been set up which allowed for an ease of generating appropriate mobility insight for COVID-19 use case and aligned with the government's foresight of initiating the project. Also, GSS being the statistical agency and lead partner was pivotal to the dissemination of the insight developed to respective government Ministries Department and Agencies (MDAs) for policy derivatives. The geographical coverage and timeliness of these analyses would not have been possible using traditional data collection methods, that is, manual surveys.
- *Principle of professional independence*: Knowing the importance of professional independence, the data pipeline developed in agreement by the partners was one that gives GSS the ability to access the aggregated, anonymized CDR data for the analyses and dissemination of mobility insight. As an integral part of the sustainability of the project, there has also been concerted efforts to boost the capacity of staff within GSS to perform these analyses which in essence would domesticate the production of appropriate statistics on mobility patterns. Also, to prevent misuse of statistics on mobility insight, a steering committee has been set up headed by the Statistician General of GSS to oversee the use, output and data dissemination.
- *Principles of privacy protection*: In the early phase of the project, issues around data security and privacy had been carefully addressed as guided by GDPR (in the absence of national guidelines) and other project specific guidelines. In addition, output analyses were expressed in percentages to the average baseline period of what was expressed as "normal" behavior which further ensured that the sensitivity of the data was protected. To ensure adherence to Ghanaian law the project partners liaised closely with the Ghana Data Protection Commission during its first phase and the Commission provided a Certificate of no Objection to the project.
- *Principles of commitment to quality*: For appropriate interpretation of the analyses, all known limitations of the CDR data used were acknowledged. For example, it is known that the dataset will only include information about people that use a Vodafone Ghana SIM card. Therefore, the entire population is not represented in the dataset, and so it is not possible to make precise statistical inferences about the whole population using only that CDR dataset. Secondly, because CDRs are an

<sup>25</sup> See <https://statsghana.gov.gh/>

actively collected telecommunications dataset—meaning that a record is generated only when a subscriber performs an action with their phone, for example, makes a call—there is data available only when subscribers are active. This will lead to situations in which, for example, a subscriber may have traveled between two regions but because they did not use their phone in both locations, their movement will not have been observed. The quality of the outputs, in particular the representativeness of the data, could be improved by onboarding more MNOs onto the initiative, as well as by incorporating data from existing surveys about phone usage, and conducting bespoke surveys. These activities require a considerable amount of time, and therefore cannot necessarily be undertaken in a rapid-response situation. However, in line with the principle of commitment to quality, the limitations of the data are noted and communicated, and improvements are planned for when time and resources are available. Details of the analysis are described in the reports that are available from the Ghana Statistical Service website. Further details of the indicators that were used in the analysis can be found on Flowminder’s COVID-19 website.<sup>26</sup>

- *Principle of international comparability:* Flowminder provided extensive documentation of its approach online during the project, which sought to facilitate international comparability. The outputs produced in response to the COVID-19 epidemic in Ghana were however only focusing on meeting the needs of the government and other stakeholders in Ghana.

**4.2.5. Data availability statement.** Access to the anonymized, aggregated CDR data, used in the work, can be requested by parties that plan to use the data for a purpose that contributes to the national development of Ghana. Requests should be directed to the Government Statistician of Ghana via e-mail: [gadmin@statsghana.gov.gh](mailto:gadmin@statsghana.gov.gh).

#### 4.2.6. Analysis methods and code

Details of the analysis are described in the reports that are available from the *Ghana Statistical Service website*. Further details of the indicators that were used in the analysis can be found on *Flowminder’s COVID-19 website*.

### 4.3. The Gambia

#### 4.3.1. Project setup<sup>27</sup>

The analysis of CDR data in The Gambia goes back to a dialog between The World Bank, the Gambia Bureau of Statistics (GBoS), and the Public Utilities Regulatory Authority (PURA) in 2019. It has explored the use of CDR data to create an evidence base for policy and project design. With the support of The World Bank and The University of Tokyo, efforts were also directed at strengthening existing data collection protocols between the MNO and the PURA. Capacity building offered an opportunity to discuss the nature of CDR data and to enhance technical capacity to produce statistics from CDR data, which helped strengthen accountability on the use of CDR data with full transparency in what data is used and how that is done.

#### 4.3.2. The objectives

The COVID-19 pandemic has highlighted the value of high-frequency and localized data to inform decision-making in a crisis situation. The spread of COVID-19 and restriction measures impacted tourism economy as well as domestic trade and economy in The Gambia. It significantly reduced urban job opportunities and economic activities, creating an urban exodus where many migrants returned to their home villages. In this context, census and survey data are reliable resources for understanding the socioeconomic aspect of affected populations, but these data alone would not be sufficient as an evidence

<sup>26</sup> See <https://covid19.flowminder.org/home>

<sup>27</sup> Further details of the project can be found in Arai et al. (2021).

base for examining the impacts of COVID-19 and following enforcement measures. To fill such data gap, this project used four types of mobility indicators as the proxies for population count, home location, distance traveled, and mobility across regions.<sup>28</sup> This use case showcases the process and outputs of the indicators produced from CDR aggregates, which are useful for understanding how population distribution and movements changed.

#### 4.3.3. *The implementation*

For COVID-19 monitoring, a set of mobility statistics are being produced from CDR data in PURA's premise and updated as new data come. CDR data are de-identified by respective MNOs to ensure that no personally identifiable information is included in the data used for producing statistics by PURA. It is used for capturing changes in population movements under COVID-19 including post-intervention periods compared with the normal routine mobility. Through the computation and analysis of mobility statistics, the following principles are taken into account:

- Existing data pipeline is used for computation, which allowed the government to quickly extend the development objective for this partnership in response to COVID-19. Although the existing data pipeline still includes some manual processes for producing statistical products, it contributed to lowering the response burden of both the regulator and MNOs while protecting the privacy of their information. It also accelerated the timeliness of government responses to data demand under COVID-19.
- Mobility indicators used for this project are not computationally intensive statistics. It does not require a lot of computational resources and skills while it can still meet the data demand to a certain extent. For example, the number of active subscribers was computed at several administrative levels at a daily basis for monitoring changes in population distributions. It could also indicate population movements over a range of temporal and spatial scales. The modal location in the evening or at night of the week were used as the proxy of home region, which could capture short-term relocation at a weekly basis.
- The preparation and analysis of CDR data are based on a protocol that is designed to eliminate data privacy issues. Any personally identifiable information was irreversibly encrypted. In addition, cell tower locations were clustered to lower spatial granularity and computed results were aggregated at the regional/district levels.
- Prior to the analysis, the validity of CDR data were examined using population density, and phone usage patterns in terms of transaction volume and number of active subscribers. Although available data are limited, this process helped understand quality in statistics produced from CDR data, and identify strengths and weaknesses to improve process and product quality.

#### 4.3.4. *The results*

Data was made available for two major MNOs, which cover approximately 70% of the market and include around 1.75 million subscribers.<sup>29</sup> The population covered by the two MNOs are primarily different socioeconomic groups. One of them is a leading MNO in The Gambia and is popular in urban areas with high-speed internet services. The other MNO provides only voice and short-messaging services with inexpensive plans, which are much popular in rural areas. It is assumed that cell phone usage in terms of transaction volumes did not change largely under COVID-19. Key results from the analysis include:

- Subscriber density highly correlates with the population density of known population data.

<sup>28</sup> Proxy for population count uses the number of unique IDs active (call, SMS, and data communication) within a day and region. Proxy for home location is defined as the modal location of the last observation on each day of the week, where the user is most frequently in the evenings or at night. For more detailed explanation please see the World Bank COVID19 Mobility Task Force repository on GitHub.

<sup>29</sup> See also <https://blogs.worldbank.org/opendata/hidden-potential-mobile-phone-data-insights-covid-19-gambia>

- Phone usage remained near-constant in terms of the number of active subscribers: The number of active subscribers is computed to examine the pattern of phone usage. This stability in the number of users suggests that people kept using their phones during the lockdown and that fluctuations in activity reflect population shifts rather than differential use patterns.
- Distribution of subscribers' home regions reflects the response to interventions and events: Result showed an increased number of people moving rural areas in response to the State of Emergency. It also reflected a spike of activity across the board, most pronounced in urban areas probably as a result of people relocating to be with their families during the holidays.

#### 4.3.5. Adherence to the principles

The use case of The Gambia demonstrated how CDR data could be leveraged for producing frequent and granular statistics to complement traditional population statistics while trying to maintain trusts in official indicators from the aspect of the five principles described in [Section 3](#).

- *Principles of necessity and proportionality.* The mobility indicators used for this project were sought to meet the data demand for monitoring changes in population movements, as well as for examining the impact of COVID-19 and interventions for future planning. Efforts were also directed to reduce the response burden of governments and MNOs by using the existing data pipeline and institutional frameworks while ensuring the privacy.
- *Principle of privacy protection.* The procedure of data pre-processing and aggregation enabled to maintain privacy. Any individually identifiable information was de-identified by respective MNOs, and de-identified data were aggregated at the regional level by PURA. Computed results include sufficient population numbers in groups, and are considered to be very hard to be re-identified. These statistics are used as an evidence base for planning and policy design, not for making decisions concerning any individual.
- *Principle of professional independence.* The methodologies employed in this initiative were examined by GBoS and PURA with their strictly professional considerations. GBoS ensured the scientific principles and professional ethics while PURA oversaw the data access and computation process to protect privacy. As part of the process to develop the data pipeline, PURA and GBoS established a platform to strengthen policy relevance of the use of CDR data. Building consensus among all stakeholders and using strategic alliances embedded in the country dialog helped foster ownership and sustainability of this initiative. These procedures including methodologies employed for computing mobility statistics are disclosed to the public with documentations and accessible at any time.
- *Principles of commitment to quality.* CDR data used for this project cover 70% of mobile subscribers, which are a subset of the national population; this is one of limitations of the data. When analysis results were interpreted, market share of the data, geographic coverage, as well as the correspondence with known population data were carefully examined to understand the validity, advantages, and weakness of data.
- *Principle of international comparability.* This information was summarized and publicly shared with the explanation on methodologies and code, which help maintain the transparency and reproductivity of the approach adopted.

4.3.6. *Data availability statement.* If possible, results of computed indicators or aggregated statistics will be made available through the website of the GBoS or the Public Utilities Regulatory Authority (PURA) (currently under the discussion).

#### 4.3.7. Analysis methods and code

Details of methodologies employed for computing indicators can be found on *the World Bank COVID19 Mobility Task Force Github repository*. Code adjusted for running a system under PURA is maintained on *the University of Tokyo's Spatial Data Commons Github repository*.

## 5. Conclusion

The COVID-19 pandemic has accelerated the use of mobile operator data to support public policy, although without a universal governance framework for its application. This article described five principles to guide and assist statistical agencies, MNOs and intermediary service providers, who are actively working on projects using mobile operator data to support governments in monitoring the effectiveness of its interventions. Compliance with each of these principles can help maintain public trust in the handling of these sensitive data and their results, and therefore keep citizen support for government policies.

Three projects (in Estonia, Ghana, and the Gambia) were described and reviewed with respect to the compliance and applicability of the five principles. Even though, in these three cases, the principles were reviewed after the completion of the projects, they still give insights into how the principles can be addressed in operational projects.

A summary of the application of the principles can be given as follows:

- *Principles of necessity and proportionality.* In all three cases mobile operator data are seen as relevant and fit-for-purpose to provide the necessary timely, frequent and geographically detailed information needed for the monitoring of human mobility. It was an advantage in Ghana and the Gambia that agreements between the NSO, the MNO, and the intermediate service providers had already been successfully completed before the start of the COVID-19 pandemic. That meant that the relevance of the use of detailed mobile operator data for compilation of statistical indicators for policy purpose had been established, including the conditions of use. The additional time also allowed the MNOs to familiarize themselves with and test the necessary algorithms to derive meaningful pre-processed and aggregated data. This was not the case in Estonia, where the urgent need for data rushed the negotiations with the MNO. Whereas the principle of privacy protection was fully implemented in Estonia, the data which were prepared and aggregated by the MNOs had to follow a simplified procedure due to lack of time to put more sophisticated procedures in place. This resulted in aggregated and anonymized data, which could be used to compile several, but not all desired indicators.
- *Principle of professional independence.* Professional independence is an essential part of the code of conduct of any NSO. Collaborations with MNOs and service providers should be based on the same principle. Transparency of methods and procedures support the professional independence. In all three use cases the NSOs worked jointly with its partners on the methods and preparation of the results. The NSOs in Ghana and Estonia also disseminated the results on their websites and could in that way exercise their professional independence.
- *Principle of privacy protection.* The protection of privacy was very similar in Ghana, the Gambia and Estonia. In all three cases the MNOs performed the data pre-processing and aggregation on the basis of procedures, which had been established with the other partners of the project. In Ghana and the Gambia these pre-processing procedures had already been tested for different purposes and these could therefore be used in the established sophisticated way also for the aggregates needed for the COVID-19 indicators. However, in Estonia there was not enough time to work out new and sophisticated pre-processing and aggregation procedures, which meant that the MNOs only pre-processed the data using simplified procedures. Nevertheless, in all three cases, privacy was well protected since the detailed CDRs stayed within the MNOs, and only anonymized and aggregated data were shared with partners outside the MNOs. In the case of Ghana, the service provider (Flowminder) was working with the MNO within the secure MNO environment. This is a model, which protects privacy and allows for more quality control at the same time.
- *Principles of commitment to quality.* For appropriate interpretation of the analyses, all known limitations of the CDR data used were acknowledged and described in the cases of Ghana and the Gambia. In the case of Estonia, the limitations of the aggregated input data were discussed. To a certain extent in all three cases, the results were analyzed against the market share of the MNO data,

the geographical coverage, as well as with respect to the coverage of the population to understand the quality of the data. Further, in line with the principle of commitment to quality, the limitations of the data are well noted and communicated, and—in all three cases—further analysis and improvements are planned for when time and resources are available.

- *Principle of international comparability.* It was not clear in the three cases exactly how international comparability of the released results could be done, given the differences in contexts and subscriber populations. However, the methodologies and code are publicly shared, which helps maintain the transparency and reproductivity of the approach adopted and could lead to develop approaches and frameworks at international level, if needed.

In conclusion, the principles of necessity and proportionality and of privacy protection are most evident and most strictly adhered to in these projects, which use mobile operator data for policy purposes. In other words, all parties in the data collaboration will make sure that the data are well protected and fit-for-purpose. Necessity and proportionality of the use of mobile operator data can be justified based on the relevance of the policy purpose (containing the spread of the virus) and the need for timely, frequent, and granular data, which allows to react in time and at local level, as needed.

The use cases demonstrated that time is needed to work out the best arrangements with the MNOs regarding the trade-off between privacy protection and the required use of the mobile operator data in terms of indicators to be derived (fit-for-purpose) and the quality checks to be made (commitment to quality). Full open access to the data would give the NSO and its service provider all possibilities to develop detailed, high-quality and relevant indicators. Fully anonymized and aggregated data limit the possibilities to develop the desired indicators. In Ghana and the Gambia these trade-off discussions had already taken place over a longer period of time for projects on different topics, but with similar requirements. In Estonia, the trade-off discussion had to be held under severe time pressure and led to the situation, in which only a simplified calculation on aggregated data could be prepared by the MNOs. The restrictions imposed on the data to protect privacy have a negative effect on the quality of the indicators, which can be compiled for policy purposes. Therefore, it is important to work out a good solution between protecting privacy and protecting quality of the indicator. For future projects, sophisticated privacy preserving techniques<sup>30</sup> can help to protect privacy while allowing more elaborate operations and more collaboration on the detailed mobile operator data (United Nations, 2019b).

Commitment to quality is an objective which the statistical agency and its partners will certainly try to achieve. However, given the time pressure and the need for early results in emergency situations, the quality standards almost necessarily have to be lowered, and difficult decisions need to be taken, if and when to release the indicators based on the assessment that the quality of the indicators is good enough. It is in general good practice, which was also applied in all three use cases, to check the new indicators against existing information before release. For example, the human mobility indicators need to be reasonable in proportion to the total number subscribers by operator per district.

Professional independence is a core principle for statistical agencies. The statistical agency must provide the policy indicators strictly according to its own professional standards also in an emergency situation while dealing with government, private sector and other parties. Whereas the methods and processes can be developed and executed jointly with the partners, the dissemination of the indicators should be strictly the responsibility of the statistical agency.

International comparability is not a priority in most cases, where a project team is trying to provide results for an emergency at national level. Nevertheless, in the project preparation stage some communication at international level may be useful, especially in a global pandemic. If international guidelines are provided to contain the spread of the virus, some effort on internationally comparable indicators may be needed as well.

<sup>30</sup> See <https://unstats.un.org/bigdata/task-teams/privacy/index.cshtml>

In all, explicitly addressing the five principles in the preparation of a project should give confidence to the statistical agency and its partners, that enough care has been exercised in the set up and implementation of the project, and should convey trust to public and government in the use mobile operator data for policy purposes.

**Supplementary Materials.** To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/dap.2021.21>.

**Data Availability Statement.** If possible, results of computed indicators or aggregated statistics will be made available through the website of the Gambia Bureau of Statistics (GBoS) or the Public Utilities Regulatory Authority (PURA) (currently under the discussion). Details of methodologies employed for computing indicators can be found on *the World Bank COVID19 Mobility Task Force Github repository*. Code adjusted for running a system under PURA is maintained on *the University of Tokyo's Spatial Data Commons Github repository*.

**Author Contributions.** Conceptualization: R.J., S.E., E.S.; Writing—original draft: R.J., K.K., S.E., E.S., K.S., L.B., T.L., W.A.A., A.A., J.N., E.M.; Writing—review and editing: R.J., S.E., E.S., W.A.A., A.A.; Supervision: R.J. All authors contributed to writing the original draft and approved the final submitted draft.

**Competing Interests.** The authors declare no competing interests exist.

**Funding Statement.** None.

#### Abbreviations

FPOS	Fundamental Principles of Official Statistics
GDPR	General Data Protection Regulation
GSMA	GSM Association
GWG	UN Global Working Group on Big Data for Official Statistics
MNO	mobile network operator
NSO	national statistical office

#### References

- Ahas R, Aasa A, Mark Ü, Pae T and Kull A (2007) Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management* 28(3), 898–910. <https://doi.org/10.1016/j.tourman.2006.05.010>.
- Arai A, Knippenberg E, Meyer M and Witayangkurn A (2021) The hidden potential of call detail records in The Gambia. *Data & Policy* 3, e9. <https://doi.org/10.1017/dap.2021.7>.
- 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. *Scientific Reports* 5, 8923. <https://doi.org/10.1038/srep08923>
- Buelens B, Daas P, Burger J, Puts M and van den Brakel J (2014) Selectivity of Big Data. *Statistics Netherlands* 2014/11.
- EDPB (2020) Guidelines 04/2020 on the Use of Location Data and Contact Tracing Tools in the Context of the COVID-19 Outbreak. Available at [https://edpb.europa.eu/sites/edpb/files/files/file1/edpb\\_guidelines\\_20200420\\_contact\\_tracing\\_covid\\_with\\_annex\\_en.pdf](https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf). (accessed 24 September 2021).
- EU (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Available at <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed 24 September 2021).
- Ghana Statistical Services, Vodafone Ghana and Flowminder Foundation (2020a) Mobility Analysis to Support the Government of Ghana in Responding to the COVID-19 Outbreak - Initial Insights into the Effect of Mobility Restrictions in Ghana Using Anonymised and Aggregated Mobile Phone Data, 3 April 2020. Available at <https://statsghana.gov.gh/COVID-19%20press%20release%20report%20-%20analysis%20overview%20-%20final1.pdf> (accessed 24 September 2021).
- Ghana Statistical Services, Vodafone Ghana and Flowminder Foundation (2020b) Mobility Analysis to Support the Government of Ghana in Responding to the COVID-19 Outbreak - Insights into the Effect of Mobility Restrictions in Ghana Using Anonymised and Aggregated Mobile Phone Data. Report #2, 15 May 2020. Available at <https://statsghana.gov.gh/COVID-19%20press%20release%20report%20-%20analysis%20overview%20-%20final.pdf> (accessed 24 September 2021).
- González M, Hidalgo C and Barabási A (2008) Understanding individual human mobility patterns. *Nature* 453, 779–782. <https://doi.org/10.1038/nature06958>.
- GSMA (2020) The GSMA COVID-19 Privacy Guidelines. Available at <https://www.gsma.com/publicpolicy/wp-content/uploads/2020/04/The-GSMA-COVID-19-Privacy-Guidelines.pdf>.

- ITU (2019) Measuring Digital Development, Facts and Figures 219. Available at <https://www.itu.int/en/mediacentre/Pages/2019-PR19.aspx>.
- Järv O, Ahas R, Saluveer E, Derudder B and Witlox F (2012) Mobile phones in a traffic flow: A geographical perspective to evening rush hour traffic analysis using call detail records. *PLoS One* 7(11), e49171. <https://doi.org/10.1371/journal.pone.0049171>.
- Jia JS, Lu X, Yuan Y, Xu G, Jia J and Christakis NA (2020) Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* 582, 389–394. <https://doi.org/10.1038/s41586-020-2284-y>.
- Kishore N, Kiang M, Engo-Monsen K, Vembar N, Schroeder A, Balsamic S and Buckee C (2020) Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. *Lancet Digital Health* 2(11), E622–E628. [https://doi.org/10.1016/S2589-7500\(20\)30193-X](https://doi.org/10.1016/S2589-7500(20)30193-X).
- Li T, Bowers R, Seidu O, Akoto-Bamfo G, Bessah D, Owusu V and Smeets L (2021) Analysis of call detail records to inform the COVID-19 response in Ghana – opportunities and challenges. *Data & Policy* 3, e11. <https://doi.org/10.1017/dap.2021.5>.
- OECD (2013) The OECD Privacy Framework. Available at [https://www.oecd.org/sti/ieconomy/oecd\\_privacy\\_framework.pdf](https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf).
- OECD (2017) Guidelines on Measuring Trust. Available at <https://www.oecd.org/governance/oecd-guidelines-on-measuring-trust-9789264278219-en.htm>.
- Saluveer E, Raun J, Tiru M, Altin L, Kroon J, Snitsarenko T, Aasa A and Silm S (2020) Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research* 81, 102895. <https://doi.org/10.1016/j.annals.2020.102895>.
- Statistics Estonia (2020a) Statistics Estonia: Mobility Has Decreased During the Emergency Situation, News Release, 3 April 2020. Available at <https://www.stat.ee/en/uudised/news-release-2020-042>.
- Statistics Estonia (2020b) Statistics Estonia: People Stay in One Location 20 Hours Per Day on Average, News Release, 9 April 2020. Available at <https://www.stat.ee/en/uudised/news-release-2020-047>.
- UNCTAD (2020) Data Protection and Privacy Legislation Worldwide. Available at [https://unctad.org/en/Pages/DTL/STI\\_and\\_ICTs/ICT4D-Legislation/eCom-Data-Protection-Laws.aspx](https://unctad.org/en/Pages/DTL/STI_and_ICTs/ICT4D-Legislation/eCom-Data-Protection-Laws.aspx) (accessed 24 September 2021).
- United Nations (2014a) Statistical Commission, Report on the Forty-Fifth Session (4–7 March 2014) E/2014/24, Decision 45/110. Available at <https://unstats.un.org/unsd/statcom/45th-session/documents/statcom-2014-45th-report-E.pdf> (accessed 24 September 2021).
- United Nations (2014b) General Assembly, Resolution Adopted by the General Assembly on 29 January 2014, 68/261, Fundamental Principles of Official Statistics. Available at <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf> (accessed 24 September 2021).
- United Nations (2019a) Handbook on the Use of Mobile Phone Data for Official Statistics. Available at <https://unstats.un.org/bigdata/task-teams/mobile-phone/> (accessed 24 September 2021).
- United Nations (2019b) Handbook on Privacy-Preserving Computation Techniques. Available at <https://marketplace.officialstatistics.org/privacy-preserving-techniques-handbook> (accessed 24 September 2021).
- Wesolowski A, Qureshi T, Boni M, Sundsøy P, Johansson M, Rasheed S, Engø-Monsen K and Buckee C (2015) Impact of human mobility on the emergence of dengue epidemics in Pakistan. *PNAS* 112(38), 11887–11892. <https://doi.org/10.1073/pnas.1504964112>.
- WHO (2020) Scientific Brief of 9 July 2020. Available at <https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions> (accessed 24 September 2021).

---

Cite this article: Jansen R, Kovacs K, Esko S, Saluveer E, Söstra K, Bengtsson L, Li T, Adewole W. A, Nester J, Arai A and Magpantay E (2021). Guiding principles to maintain public trust in the use of mobile operator data for policy purposes. *Data & Policy*, 3: e24. doi:10.1017/dap.2021.21