

Design & Implementation of Mobile Phone Data Initiatives:

A Practical Manual

A UN-CEBD Mobile Phone
Data Task Team Handbook



© 2026 The authors

This work is prepared for the UN-CEBD Mobile Phone Data Task Team.

Publisher: United Nations Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team.

Except where otherwise noted, this work is licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0). This licence does not apply to UN/CEBD logos, trademarks, third-party images, or other third-party material reproduced with permission. Reuse of the manual must not imply endorsement by the authors, the United Nations, UN-CEBD, or affiliated organisations.

DOI: <https://doi.org/10.5281/zenodo.20498954>

Suggested citation:

Riley, C., Rowe, F., Magpantay, E., Eyre, R., Delaporte, S., Harrison, J., Lefebvre, V., Smallwood, T., Chaves, L., Ruiz, P., Sales, M. H., Picornell, M., Rüütli, E., Vent, K., Esko, S., Saluveer, E., Arai, A., Blanchard, P., Milusheva, S., and Monroe, T. (2026). *Design and Implementation of Mobile Phone Data Initiatives: A Practical Manual*. United Nations Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team. CC BY 4.0. <https://doi.org/10.5281/zenodo.20498954>. <https://un-cebd-mobile-data.github.io/mobile-phone-data-handbook/>

Digital material and resources associated with this manual are available at <https://un-cebd-mobile-data.github.io/mobile-phone-data-handbook/>

Table of contents

Preface	7
Authors	7
Recommended citation	8
Licence	8
Acknowledgements	9
Authors and contributors	9
Funders	10
Glossary of Terms and Abbreviations	11
1 Chapter 1: Planning a Mobile Phone Data Initiative	20
1.1 Understanding the Basics of MPD	21
1.1.1 What Is MPD?	21
1.1.2 MNO data	21
1.1.3 The spatial and temporal characteristics of CDRs vs Signalling and GPS data	23
1.2 Clarifying the Value and Purpose of Using MPD	25
1.2.1 Strategic Value and Use Cases	25
1.2.2 The Importance of a Clear Purpose	25
1.2.3 Developing a Theory of Change	26
1.3 Core Technical and Institutional Requirements	27
1.3.1 Data Access Arrangements	27
1.3.2 Infrastructure and Data Processing Capacity	27
1.3.3 Human Resources and Leadership	27
1.4 Legal, Regulatory, and Ethical Context	29
1.5 Stakeholders Ecosystem and Governance	30
1.6 Risk management and mitigation measures	32
1.6.1 Data-related risks	32
1.6.2 Risk mitigation measures	33
1.7 Designing for Long-Term Sustainability	34
1.8 Conclusion	34
2 Chapter 2: Policy Applications for MPD	35
2.1 Why and How is MPD Useful?	35
2.2 Information Society Statistics	36
2.2.1 Information Society and Digital Inclusion	36

2.2.2	Network Coverage and Infrastructure Planning	37
2.3	Population Statistics and Mobility Analysis	37
2.3.1	Dynamic Population Mapping and Migration	37
2.3.2	Disaster Management	38
2.3.3	Public Health and Epidemiology	40
2.3.4	Tourism Statistics	41
2.3.5	Transport and Mobility Planning	43
2.4	Socio-Economic Applications: Poverty Mapping	44
2.4.1	Bangladesh poverty mapping study	45
2.4.2	Togo’s use of MPD for targeting social protection payments	45
2.4.3	Testing the utility of MPD for poverty prediction in Ghana	45
2.5	Conclusion	46
3	Chapter 3: Arranging Partnerships and Data Access	47
3.1	Understanding the Central Role of MNOs	47
3.2	Why Operators May Hesitate to Share Data	48
3.3	Managing and Mitigating MNOs’ Risks	48
3.4	Articulating Clear Incentives for MNO Participation	49
3.5	Defining Boundaries Between Public and Private Data Use	50
3.6	Understanding Data, Technical, and Legal Roles	50
3.7	Using Maturity Assessment to Select a Data Access Model	51
3.8	Formalising Partnerships Through Agreements	51
3.9	Conclusion	52
4	Chapter 4: Data Processing and Data Pipelines for MPD Initiatives	53
4.1	Overview of the MPD Pipeline	53
4.2	Data Collection and Generation	53
4.3	Extract, Transform, and Load (ETL)	54
4.4	Data Cleaning: Ensuring Analytical Validity	54
4.5	Core Data Processing and Methodological Models	55
4.6	Aggregation and Indicator Construction	56
4.7	Scaling, Bias Adjustment, and Quality Assurance	56
4.8	Privacy by Design in the Pipeline	56
4.9	Pipeline Deployment Models	57
4.10	Conclusion	57
5	Chapter 5: Data Quality and Characteristics	58
5.1	How CDRs are generated and key data fields	58
5.1.1	Pseudonymisation and identifiers: why they matter for quality	59
5.2	Strengths and limitations of CDR data	59
5.2.1	Strengths: coverage, timeliness, and passive generation	59
5.2.2	Limitations: spatial precision, temporal intermittency, and representativeness	59
5.3	Interpreting location in CDR: from records to meaningful places	60
5.3.1	Mobility as “movement between towers,” not continuous location traces	60

5.3.2	Using timestamps to infer home and work	60
5.4	Spatial coverage and the Voronoi approximation: what it enables and what it hides	61
5.4.1	Voronoi cells as a workable model of tower coverage	61
5.4.2	Why CDR trajectories can deviate from “actual mobility”	61
5.5	Key data-quality considerations driven by network behaviour	61
5.5.1	Tower density and the visibility of short trips	61
5.5.2	Handover and location noise: false mobility signals	62
5.6	Representativeness and bias: why “who is in the data” is central to quality	62
5.7	Adjusting for bias: population-weighted adjustment and triangulation with other sources	63
5.8	Quality assurance and trust: moving from analysis to statistics suitable for policy use	64
5.8.1	Why quality assurance is not optional	64
5.8.2	A structured quality assurance approach: ESS QAF (European Statistical System 2019) and “quality gates”	64
5.9	Conclusion	65
6	Chapter 6: Data Governance and Safeguards in MPD Initiatives	66
6.1	What Does It Mean to Process Data?	66
6.2	Data Governance and the Stakeholder Ecosystem	67
6.3	Personal and Non-Personal Data in MPD Initiatives	67
6.4	Legal and Regulatory Context	68
6.5	Risks Associated with MPD Initiatives	68
6.6	Safeguards and Mitigation Measures	68
6.7	Best Practices for Protecting Privacy in CDR analysis	69
6.7.1	Current, standard methods for preserving individual privacy	69
6.7.2	Emerging and novel Privacy-Enhancing Technologies (PETs)	70
6.8	International frameworks and industry standards	71
6.8.1	UN Guiding Principles for maintaining public trust when using MPD	71
6.8.2	The Locus Charter	72
6.8.3	GSMA Mobile Privacy Principles	73
6.9	Conclusion	74
7	Chapter 7: Managing the Communications Aspects of MPD Initiatives	75
7.1	Foundational Communication Principles	75
7.1.1	Knowing and Segmenting the Audience	75
7.1.2	Avoiding Jargon and Explaining Acronyms	76
7.1.3	Starting with the “Why” and the Public Value	76
7.1.4	Transparency and Honesty	77
7.1.5	Allocating Dedicated Communication Resources	77
7.2	Communication Principles Specific to MPD Initiatives	77
7.2.1	Clarifying What Is Sensitive	77
7.2.2	Explaining Data Sources and Safeguards Accessibly	78
7.2.3	Using Careful and Responsible Language	78
7.2.4	Reiterating the Purpose and Benefits	78

7.3	Transparency, Independence, and Public Engagement in Statistical Contexts	78
7.4	Balancing Rigour, Accessibility, and Uncertainty	79
7.5	Managing Communications Within Partnerships	80
7.6	Conclusion	80
Appendices		81
A Appendix 1: Further recommended resources		81
A.1	Chapter 1: Planning a Mobile Phone Data Initiative	81
A.2	Chapter 2: Policy Applications	82
A.3	Chapter 3: Arranging Partnerships and Data Access	83
A.4	Chapter 4: Data Processing and Data Pipelines	83
A.5	Chapter 5: Data Quality and Characteristics	84
A.6	Chapter 6: Data Governance and Safeguards	85
References		86

Preface

Project status

Publisher: United Nations Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team. DOI: [10.5281/zenodo.20498954](https://doi.org/10.5281/zenodo.20498954). Except where otherwise noted, this work is licensed under a [Creative Commons Attribution 4.0 International Licence \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). The public website is configured for GitHub Pages at <https://un-cebd-mobile-data.github.io/mobile-phone-data-handbook/>.

This handbook provides an in-depth guide to planning and sustaining a Mobile Phone Data (MPD) initiative, with a primary focus on the use of Call Detail Records (CDRs) for public policy, statistical, and development purposes, including operational decision-making. It builds on, and develops further, the concepts and principles first described in the original Handbook on the Use of Mobile Phone Data for Official Statistics released by what was then known as the UN Global Working Group on Big Data for Official Statistics. ([United Nations Statistics Division 2019](#))

The handbook is intended for practitioners working in national statistical offices, telecom regulators, mobile network operators, government ministries, and partner organisations who would like to initiate an MPD initiative. It also contains advice and guidance for those who may already have embarked on the journey of establishing such an initiative but who are searching for more information or guidance on how to do so effectively and sustainably. It is designed to enable such readers to understand not only the steps involved in planning an MPD initiative, but also the technical, institutional, legal, and ethical reasoning that underpins each decision. It is suitable for both technical and non-technical audiences, and does not assume deep prior technical expertise in MPD analytics.

Authors

Cathy Riley, Francisco Rowe, Esperanza Magpantay, Robert Eyre, Sophie Delaporte, James Harrison, Veronique Lefebvre, Thomas Smallwood, Luisa Chaves, Pablo Ruiz, Maria Henar Sales, Miguel Picornell, Egle Rütli, Kaisa Vent, Siim Esko, Erki Saluveer, Ayumi Arai, Paul Blanchard, Sveta Milusheva, and Trevor Monroe.

Recommended citation

Note

Riley, C., Rowe, F., Magpantay, E., Eyre, R., Delaporte, S., Harrison, J., Lefebvre, V., Smallwood, T., Chaves, L., Ruiz, P., Sales, M. H., Picornell, M., Rütli, E., Vent, K., Esko, S., Saluveer, E., Arai, A., Blanchard, P., Milusheva, S., & Monroe, T. (2026). *Design and Implementation of Mobile Phone Data Initiatives: A Practical Manual*. United Nations Committee of Experts on Big Data and Data Science for Official Statistics, Mobile Phone Data Task Team. CC BY 4.0. <https://doi.org/10.5281/zenodo.20498954>. <https://un-cebd-mobile-data.github.io/mobile-phone-data-handbook/>

Licence

Except where otherwise noted, the text of this manual is licensed under a [Creative Commons Attribution 4.0 International Licence \(CC BY 4.0\)](#).

This licence does not apply to UN/CEBD logos, trademarks, third-party images, or other third-party material reproduced with permission. Reuse of the manual must not imply endorsement by the authors, the United Nations, UN-CEBD, or affiliated organisations.

Acknowledgements

This manual is the product of collective effort by organisations and individuals working on mobile phone data. It was prepared under the auspices of the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) Task Team on Mobile Phone Data, chaired by Esperanza Magpantay (ITU), and vice-chaired by Cathy Riley (Flowminder Foundation) and Francisco Rowe (University of Liverpool).

The content draws extensively on inputs provided by active members of the UN Committee of Experts on Big Data and Official Statistics (UN-CEBD) Mobile Phone Data Task Team (MPD-TT). It also draws extensively from materials and experiences generated through the ITU/World Bank Global Data Facility Mobile Phone Data (GDF-MPD) project, whose resources and capacity building efforts provided an important foundation for many of the approaches, examples, and practices presented herein.

Authors and contributors

The lead author and editor of this handbook was Cathy Riley, Strategic Partnerships Director at Flowminder Foundation and Project Director of the “Foundations in Mobile Phone Data for Policy and Statistics” training project. Co-authors of content presented in this manual, and their organisational affiliations, are as follows:

- **Flowminder Foundation:** Robert Eyre, Sophie Delaporte, James Harrison, Veronique Lefebvre, Thomas Smallwood
- **International Telecommunication Union (ITU):** Esperanza Magpantay
- **Nommon:** Luisa Chaves, Pablo Ruiz, Maria Henar Sales, Miguel Picornell
- **Positium:** Egle Rüütli, Kaisa Vent, Siim Esko, Erki Saluveer
- **University of Liverpool:** Francisco Rowe
- **University of Tokyo:** Ayumi Arai
- **World Bank:** Paul Blanchard, Sveta Milusheva, Trevor Monroe

Additional contributions, including comments on drafts and content for case studies, were received from Titi Kanti Lestari, Matthias Offermans and Avik Sarkar. The formatting and development of the e-book version of the Manual were led by Francisco Rowe (University of Liverpool), whose efforts contributed substantially to improving the accessibility and presentation of the publication.

Funders

The compilation of this manual was made possible through in-kind contributions from most of the organisations involved in producing it. Flowminder Foundation wishes to acknowledge the contribution of funds from the UK Foreign, Commonwealth and Development Office (FCDO) towards enabling staff attendance at the UN-CEBD MPD-TT meetings necessary for coordination efforts in production of this handbook.

The original content from which this handbook has drawn substantially was funded by the World Bank's GDF-MPD window, a multi-partner window funded by the Spanish Government, Ministry of Economy, Commerce and Business and jointly managed by DEC (DECDG, DIME), Poverty and Digital Development Global Practices in technical partnership with the International Telecommunication Union (ITU).

Glossary of Terms and Abbreviations

Acronym	Term	Synonyms	Definition
MPD	Mobile Phone Data	Mobile Positioning Data, Mobile Network Data, CDR data, Signalling data, MNO data	A term used to describe the data source representing location data from the logs of Mobile Network Operators (MNOs)
	Local Administrative Units	Administrative Division	Subnational administrative divisions, such as municipalities or districts, used for governance and statistical reporting
	Point of Interest		A specific location of significance, such as a landmark, business, or facility, used in MPD processing to analyse mobility patterns and contextualise user behaviour within geographic spaces
	Raw data	Raw MNO data	Raw CDR/DDR/XDR/Signalling data from a mobile network operator (MNO)
	Pseudonymised data		Identifier hashed and only relevant fields remain

Acronym	Term	Synonyms	Definition
	Pre-processed data		Cleaned and pre-processed data. Can include some aggregation (hour and approximate location)
	Processed data		Processed MPD data into continuity model (stay and move sections)
	Individual aggregation		Individually aggregated data output (e.g. individual anchors)
	Aggregates		Aggregates according to business rules for quality checking (e.g. statistical indicators)
	Statistical indicators	Indicators	Statistical indicators that have undergone QA and statistical disclosure control
CDR	Call Detail Records		Mobile network operator event data about the call activities (calling, sending SMS etc). Usually within billing database

Acronym	Term	Synonyms	Definition
IPDR	Internet Protocol Detail Records	Data Detail Records (DDR, XDR)	Mobile network operator event data about the usage of the internet. These can be useful data fields to request from MNO billing records as they capture both active and passive internet engagement (e.g. background app data usage) by subscribers.
	Signalling data	Probe data	Signalling data is the extra data generated in communication networks to support users in having a smooth connection. Usually includes everything from xDR to location updates, cell handover to other measurement reports at the base station level. This type of data is frequently generated in the network, but usually deleted after a short period of time. Separate network probes are installed to store the data, which results in very voluminous datasets.

Acronym	Term	Synonyms	Definition
	Domestic MPD	Domestic data	Domestic subscribers data in the country's operator's database
	Inbound MPD	Inbound data, Roaming in MPD	Foreign mobile network operator subscribers roaming in the country's operator's database
	Outbound MPD	Outbound data, Roaming Out MPD	Domestic mobile network operator subscribers roaming outside the country (data with a delay in the country's operator's database)
	Event Data	Records	Mobile network generated records, such as CDRs or location updates, that capture discrete user activities (time and location) and are main component of MPD
	Network Data		Information about the mobile network infrastructure, including cell tower locations, coverage areas, and network configurations, used to interpret and process MPD for spatial and mobility analysis

Acronym	Term	Synonyms	Definition
CRM	Customer Relationship Management Data Reference Data	Client Portfolio Data	Information that the operator has for each contract. It typically includes sociodemographic attributes such as age, gender, place of residence, and occupation, with varying levels of accuracy and completeness. Additional datasets, such as surveys, census, or official statistics used as a benchmark to validate and assess the quality and representativeness of MPD.
QA	Quality Assurance		A systematic process ensuring the accuracy, consistency, and reliability of MPD
ETL	Extract, Transform, Load		A data processing workflow that involves extracting data from sources, transforming it into a suitable format, and loading it into a target system
VPN	Virtual Private Network		A secure network connection that encrypts data transmission and allows remote access to private networks over the internet

Acronym	Term	Synonyms	Definition
	Country of Residence		The country where a mobile network operator subscriber has their primary long-term presence
	Place of Residence	Home location, Home anchor	The main home location where a mobile network operator subscriber has their primary long-term presence
	Home Detection Algorithm	Anchor Point methodology	A method used to determine a mobile subscriber's primary residence based on long-term mobile network usage patterns
	Stay		A period when a person remains within a limited area for a certain duration.
	Movement	Move	A transition between two distinct locations (between 2 consecutive stays).
	Usual Environment		The area where an individual regularly resides and carries out daily activities
OD matrices	Origin Destination matrices		Statistical tables that represent the movement of individuals between different geographic locations
SDC	Statistical Disclosure Control		Techniques applied to protect individual privacy in statistical data by preventing the identification of personal information

Acronym	Term	Synonyms	Definition
BTS	Base Transceiver Station	Cell	Base Transceiver Station (same as mobile network cell, or simply cell)
CGI	Cell Global Identity	Full cell identifier	A unique number that identifies every subscriber on a cellular network, stored on the SIM card. Composed of the Mobile Country Code (MCC), Mobile Network Code (MNC), and Mobile Subscription Identification Number (MSIN). Used by networks to authenticate subscribers and facilitate roaming.
	Cell Coverage Area		A specific zone defined for each BTS, which includes the whole area where a mobile subscriber can be served by the BTS.
	Cell Tower	Cell base tower, Cellphone tower, Tower	A single location housing multiple BTSs. Typically, a cell tower consists of a radio mast, tower, or other elevated structure that supports several BTSs (ranging from a few to a few dozen) positioned at different heights and orientations to provide coverage in all directions.

Acronym	Term	Synonyms	Definition
	Cell Tower Coverage Area		A specific zone defined for each Cell Tower, which includes the whole area where a mobile subscriber can be served by at least one BTS contained in the Cell Tower.
	Mobile subscriber	User	Person registered and actively using a mobile service from a mobile network operator.
IMSI	International mobile subscriber identity		International mobile subscriber identity. Associated with a SIM card. Consists of MCC, MNC and subscriber identity number.
IMEI	International Mobile Equipment Identity		International Mobile Equipment Identity. Associated with a device/handset. Helpful to filter out IoT devices.
MCC	Mobile Country Code		A three-digit numeric code, defined by ITU-T Recommendation E.212, used to identify the country associated with a mobile network. Used in combination with the Mobile Network Code (MNC) to uniquely identify a mobile network operator, and as a component of the IMSI.

Acronym	Term	Synonyms	Definition
MSISDN	Mobile Station International Subscriber Directory Number		The phone number assigned to a mobile subscriber that is used for making and receiving calls
MNO	Mobile Network Operator		A company that provides mobile telecommunications services, including infrastructure, spectrum, and customer subscriptions
NSO	National Statistic Office		A government agency responsible for collecting, analysing, and disseminating official statistics
TRA	Telecommunications Regulatory Authority		The national body responsible for regulating and overseeing telecommunications services and policies
DPA	Data Protection Authority		A regulatory body responsible for enforcing data protection laws and ensuring privacy rights are upheld
ITU	International Telecommunication Union		A United Nations agency that coordinates global telecommunication and ICT standards and policies

1 Chapter 1: Planning a Mobile Phone Data Initiative

This chapter introduces the foundational considerations for planning a Mobile Phone Data (MPD) initiative, setting the stage for more detailed technical, institutional, and operational guidance in subsequent chapters. As mobile phones have become nearly ubiquitous across diverse socioeconomic contexts, the digital traces they generate offer unprecedented opportunities to inform public policy, development planning, and humanitarian action (Rowe 2022; Rowe et al. 2023). Harnessing this potential, however, requires careful planning, strong governance, and a clear understanding of both the opportunities and constraints associated with MPD.

The chapter begins by defining what is meant by MPD and outlining the principal ways in which it can be used to generate insights on population dynamics, mobility patterns, service access, and behavioural trends. It then examines the motivations for undertaking an MPD initiative, including the policy and operational gaps such initiatives can address, as well as the comparative advantages of MPD relative to more traditional data sources. Building on this foundation, the chapter identifies the key technical, legal, institutional, and organisational requirements that must be in place to ensure an MPD initiative is effective, ethical, and fit for purpose.

Recognising that MPD initiatives are inherently multi-actor endeavours, the chapter also discusses the critical stakeholders who must be engaged throughout the planning process, clarifying their respective roles, responsibilities, and incentives. It further highlights the principal risks associated with MPD initiatives, such as privacy concerns, data misuse, capacity constraints, and sustainability challenges, and outlines how these risks can be anticipated and mitigated through proactive planning. The chapter concludes by presenting approaches for assessing a country's readiness to implement MPD initiatives and by introducing key design principles to support long-term sustainability, institutionalisation, and impact. Together, these elements provide a structured framework for decision-makers and practitioners seeking to responsibly and effectively integrate MPD into their data ecosystems.

1.1 Understanding the Basics of MPD

1.1.1 What Is MPD?

MPD refers to digital traces generated through the operation and use of mobile communication devices ([Rowe and González-Leonardo 2024](#)). These traces are created as mobile phones interact either with mobile network infrastructure or with software applications installed on the device. Across all forms, MPD has one defining characteristic: it can be used to approximate the geographic position of a device, and by extension its user, over time. This makes it particularly valuable for analysing patterns of human mobility and population dynamics ([Gonzalez et al. 2008](#); [Song et al. 2010](#); [Blondel et al. 2015](#)).

There are two broad categories of MPD:

1. **Mobile Network Operator (MNO) data**, which is generated within the telecommunications network itself as part of routine service provision, such as CDRs and signalling data; and
2. **GPS-based data**, which is generated by smartphones and collected by applications when users opt in.

This training manual focuses exclusively on the former, and more specifically on CDRs, because this is the data type most commonly available at national scale and most frequently used in official statistics and public policy applications ([United Nations Statistics Division 2019](#); [Ricciato et al. 2020](#); [Salgado et al. 2021](#)).

1.1.2 MNO data

MNO data refers to data that is generated and stored by the telecommunications companies who operate the infrastructure such as cell tower networks which power mobile phone communications. These companies collect and store data about the operation of their systems for different reasons and purposes. What data they collect and store depends to some extent on various factors including the particular company's internal policies, operating procedures, licensing requirements and available infrastructure (especially for storing large amounts of big data). We discuss here two main types of MNO data: CDRs and signalling data.

1.1.2.1 CDRs

CDRs are generated as part of the process for billing customers. They are transaction logs created by MNOs whenever a subscriber uses their network services. These services include making or receiving a voice call, sending or receiving an SMS, or using their phone for access to the internet such as when downloading emails, searching websites or engaging with social media (this type of event is called a mobile data session). CDRs are generated passively and automatically; no additional action is required from the user beyond normal phone usage. This passive generation is one of the benefits of CDRs - they don't incur the costs and delays associated with field data collection.

From a technical perspective, a standard CDR contains a limited but highly structured set of variables. These typically include a subscriber identifier, a cell tower or cell sector identifier indicating the network element that handled the communication, a timestamp marking when the event occurred, and a code describing the type of network event. Additional fields may include an identifier for the receiving party or technical routing information, depending on the operator and the use case. It is important when third parties are accessing the CDR data to ensure that any subscriber identifiers have been Pseudonymised (see section 5.1.2 for details).

It is critically important to understand what CDRs do *not* contain. They never include the content of calls, messages, or internet activity. Analysts cannot see what was said, written, or accessed online. This distinction is central to both ethical communication and legal compliance, and it should be clearly articulated to stakeholders and the public. For more on how to communicate about MPD initiatives see Chapter 7.

1.1.2.2 Signalling Data

Signalling data (sometimes called network signalling, probe data, or cellular signalling traffic) is the stream of technical “keep the network working” events generated by continuous communication between a handset and the mobile network. These events are not only generated when a person makes a call/SMS or starts a data session, but they also triggered by mobility and location-management mechanisms, for example as a subscriber moves between cell towers the system handovers that occur as the device moves between cells will generate signals, so will the “attach or detach” activity necessitated by a phone powering on or off. In practice, the records look similar to CDR data but the dataset is much bigger because the event universe is much larger than billing-oriented datasets due to the fact that it includes operational signalling generated by the network itself and by idle devices, not just data produced when subscribers actively use their phones. This is both a benefit (more data is available) and a practical challenge, because signalling datasets can be enormous, making data storage, access governance, and processing materially harder than is the case for CDR pipelines.

1.1.2.3 Mobile Phone GPS-derived Data

GPS-derived MPD refers to location information captured directly by the **global positioning system (GPS) sensors** embedded in smartphones and other mobile devices, typically via apps that have permission to record and share location (Barreras and Watts 2024). Unlike the network-generated datasets described above, GPS data are collected from a device's onboard navigation chipset and can provide **latitude/longitude coordinates, with high geographical precision (often within a few metres)** and **temporal frequency** (e.g., seconds to minutes). They can collect continuous data on devices anywhere in the globe, thus offering global data coverage. However, the geographical precision and temporal frequency of data may vary depending on how the app is configured, users engage with the app, user consent is managed, and the type of technology used to build the device collecting the data.

Overall, GPS-derived smartphone datasets are a powerful complement to network-generated data in mobility analytics. They trade **population coverage** for **precision, detail and geographical coverage** in the recorded spatial trajectories. The geographical coverage enables more easily capturing cross-national movements. The use of CDR data to capture such moves is more challenging as people may switch SIM cards, mobile phone devices or / and operators, preventing the recording of an interrupted sequence of device's locations. Thus, the choice between GPS data, or their integration with CDR data, depends on the data and analytical specification requirements of the task at hand, such as whether **fine-grained individual movement paths** or **broad population flow patterns** are more central to the application.

1.1.3 The spatial and temporal characteristics of CDRs vs Signalling and GPS data

1.1.3.1 The spatio-temporal characteristics of CDR Data

The spatial resolution of CDR data is determined by the mobile network infrastructure rather than the device itself. Location is inferred from the cell site providing the service. This acts as a proxy for the user's position, capturing the interaction between the mobile network infrastructure and geographic position of devices. In dense urban environments, cell towers may cover relatively small areas, resulting in finer spatial granularity. In rural or remote areas, a single tower may cover a larger area, leading to coarser location estimates. Planners must account for this variability when assessing whether CDRs are suitable for a particular analytical purpose and application (Blondel et al. 2015; Ricciato et al. 2020).

The frequency of events may affect the temporal resolution of CDR data. Data are generated based on the occurrence of events reflecting user behaviour, rather than being a continuous data stream per se (other forms of higher resolution data from MNOs include what may be called signalling or ping data). In CDRs, a user's location can only be estimated when the user actively uses the network. As a result, the temporal density of CDRs can vary widely across individuals and contexts, influenced by factors such as phone ownership, usage patterns, socioeconomic status and network pricing. This intermittency introduces analytical challenges that must be addressed through appropriate statistical methods, data integration and careful interpretation (Blondel et al. 2015; Wesolowski et al. 2013; Ricciato et al. 2020).

1.1.3.2 Comparison of CDRs to signalling and GPS-derived data

Relative to CDRs, the key difference in signalling data is sampling frequency and who gets observed. CDRs are typically generated when a billable telecommunications event occurs (call/SMS/data session), so people who use their phones infrequently can appear "missing" for long stretches of time. Signalling data can produce observations every few seconds/minutes, while the phone is switched on, including for passive/idle devices, which often yields much higher temporal granularity and better continuity for mobility mapping; especially for inferring trips, dwell times and flows for low-usage subscribers.

Alternatively, GPS data tends to have a much higher spatial precision and geographical coverage (as opposed to population coverage) than network-generated data. That is because the location information originates from satellite signals interpreted by the phone's operating system and applications. However, GPS datasets can be skewed by aspects such as user opt-in, app usage patterns and battery-saving behaviours, since data collection typically requires explicit permission and can be turned off or throttled by users or the OS. Additionally, GPS datasets may provide greater geographical coverage offering an opportunity to capture cross-national movements, but they generally provide more limited population coverage than CDR data. A single MNO may cover over half of the mobile phone user population, whereas GPS data from a single application, such as AirBnB may represent a much smaller share of mobile phone users, particularly seeking accommodation.

Additionally, GPS data may be susceptible to additional sources of biases than CDR and signalling data. GPS data are generated primarily from *smartphones* reducing the potential sample size of data collection and recording from a selective, likely wealthier segment of the population. This may be particularly true in low-income economies. By contrast, CDR and signalling data cover both smartphones and feature phones, often achieving far higher population coverage. In many low- and middle-income contexts, and depending on the use case, the broader coverage of CDRs can outweigh the lower precision available through such datasets, particularly for national-level analysis and policy monitoring. Duplication may thus create additional representativeness biases in GPS data. GPS data may also contain duplicate information generated from the same device but captured through two or more applications, creating further distortions in the data.

In summary: Compared to GPS app data, CDRs and signalling data usually have lower spatial precision (often cell/sector-level rather than meter-level) and geographical coverage but they can have broader population coverage because it is network-side rather than opt-in/app-instrumented. In contrast, GPS tends to be higher resolution and more geographically precise potentially at the expense of smaller population coverage and a larger number of sources of bias.

1.2 Clarifying the Value and Purpose of Using MPD

1.2.1 Strategic Value and Use Cases

The growing interest in MPD stems from its ability to complement traditional data sources such as censuses, household surveys, and administrative records. These traditional sources are indispensable but often expensive, infrequent, and slow to update. MPD, by contrast, is generated continuously and can provide near real-time insights. ([Blondel et al. 2015](#)) Typical motivations for launching an MPD initiative include:

- Addressing temporal or spatial gaps in existing statistics
- Enhancing the timeliness of policy-relevant indicators
- Supporting rapid decision-making during crises or shocks ([Rowe et al. 2023](#))
- Improving understanding of population mobility and service access

Each of these motivations carries different technical and governance implications. For example, crisis response applications may prioritise speed and automation, while official statistics may emphasise methodological rigor and reproducibility. Different policy and statistical applications are presented in more detail in Chapter 2.

1.2.2 The Importance of a Clear Purpose

A recurring lesson from past initiatives is the central importance of clearly defining the purpose of data access from the outset. MPD is powerful, but it is not universally appropriate. Planners should explicitly ask whether the policy or analytical question at hand genuinely requires MPD, or whether it could be answered using simpler, less sensitive, or less costly data sources.

A clearly articulated purpose guides decisions about which variables are needed, how frequently data should be updated, which stakeholders must be involved, and what level of investment is justified. Without this clarity, initiatives risk becoming technically complex without delivering commensurate public value.

1.2.3 Developing a Theory of Change

Developing a Theory of Change can be helpful when planning MPD initiatives. The idea of Theory of Change is it provides a practical planning framework for articulating how the activities being undertaken are expected to lead to meaningful policy and development impacts. Rather than relying on implicit assumptions or linear thinking, the idea is to encourage those planning an MPD project or programme to map out the full causal pathway from what they do to what they want to achieve, making explicit the links between activities, outputs, outcomes, and impact.

Box 1: Developing a Theory of Change: a Tool for Planning based on Purpose In the context of MPD initiatives, producing a Theory of Change is a particularly valuable undertaking because these initiatives operate within complex data ecosystems involving multiple stakeholders, technical systems, legal constraints, and institutional incentives. The framework can help initiative designers move beyond the assumption that simply accessing MPD will automatically improve decision-making, and instead examine what must happen at each stage for data to be used effectively and responsibly.

A typical Theory of Change begins by clarifying the intended impact, such as improved evidence-based policymaking, sustainable integration of MPD into national data systems, responsible data use, or enhanced ability to monitor and anticipate policy challenges. From there, it works backward to identify the outcomes that must be in place for that impact to occur, the outputs required to achieve those outcomes, and the specific inputs and activities needed to produce those outputs.

Crucially, Theory of Change requires explicit identification of preconditions and assumptions at each step. In MPD projects these can include aspects related to: stakeholder commitment, institutional leadership, legal and regulatory feasibility, technical capacity, data quality, staff incentives, and availability of complementary data sources. Being clear about the Theory of Change also draws attention to contextual factors such as political, institutional, or environmental considerations that may enable or hinder progress. By identifying “killer assumptions” during early stages of design and planning, initiatives can adjust plans, mitigate risks, and recognise where further analysis may be needed.

For example, a seemingly simple sequence of events, such as partners signing a data-sharing agreement document in which they agree to collaborate on receiving data and conducting analysis, will be dependent on multiple underlying conditions. These include trust between partners, adequate data governance, secure and privacy-protecting data pipelines, staff skills and availability to do analysis, and the effective engagement and alignment of all relevant stakeholders. Theory of Change provides a structured way to test whether these conditions are realistically in place.

Applying Theory of Change to MPD initiatives is recommended because it can help practitioners design more robust, context-sensitive projects. It strengthens planning by clarifying causal logic, improves the likelihood that activities will lead to desired outcomes, and supports learning by making assumptions explicit and open to review.

1.3 Core Technical and Institutional Requirements

1.3.1 Data Access Arrangements

Access to high-quality CDR data is the foundational requirement of any MPD initiative. In practice, this access is secured through formal agreements with MNOs or, in some contexts, telecom regulators. These agreements must specify the scope of data shared, the frequency of updates, permitted uses, retention periods, and security requirements.

From a technical standpoint, planners must ensure that the data provided includes not only CDR event logs but also the associated data required for interpretation, particularly cell tower location information. These datasets need to be accurate and up-to-date, otherwise the subsequent analytical results may be misleading or invalid (see also Chapter 5 for more information on data characteristics and quality).

1.3.2 Infrastructure and Data Processing Capacity

CDR datasets are typically large, often comprising millions or billions of records per day. Processing such volumes requires adequate computational infrastructure, including secure servers, scalable storage, and efficient data processing frameworks. Decisions must be made regarding whether this infrastructure is hosted within a government environment, provided by an MNO, or operated through a trusted third party. Software choices also matter. While many analytical tasks can be performed using open-source tools, the organisations involved must ensure that staff have the skills to use them effectively and securely.

1.3.3 Human Resources and Leadership

Technical infrastructure alone is insufficient for an MPD initiative without the involvement of skilled personnel, both technical and non-technical. MPD initiatives typically require a multidisciplinary team, including data engineers to manage ingestion and pipelines, analysts and data scientists to develop and implement the relevant indicators and models, and project managers to coordinate delivery, liaise with stakeholders, mobilise resources and monitor timelines. Furthermore, experience shows that strong, clearly designated leadership is essential. Champions who will drive forward the initiative and coordinate the relevant role players play an absolutely critical role in maintaining strategic focus, resolving institutional tensions, and ensuring that technical work remains aligned with the initiative's ultimate objectives.

Typically the functions that should be considered for an effective initiative include:

- **Project management:** Individuals who will serve as the overall coordinator of work plans and be the key point of contact for other stakeholders will be required, as a minimum, at both the MNO and the end-user organisation (e.g. the national statistical office, or a Ministry). These roles help to ensure that the initiative is well managed, runs smoothly and meets its objectives by enabling other technical staff to focus on their components in a coordinated fashion. Effective project managers in an MPD context must be comfortable operating across institutional boundaries, translating technical constraints into operational implications, managing risks related to timelines and data access, and ensuring that governance and approval processes are respected. They are also responsible for documenting decisions, managing scope, and establishing clear reporting lines.
- **Data engineering:** Any MPD initiative will require the expertise of IT, infrastructure or data engineers who will handle obtaining access to the data (which can include, for instance, setting up secure VPNs), getting the data into the necessary structures and formats to enable analysis to take place. MNO staff working with billing records and network data will need to be involved. In order to effectively engage with the technical staff at the MNO, the receiving organisation will also need to have staff with knowledge and capabilities associated with such skills. Data engineers are responsible for designing and maintaining secure data transfer mechanisms, implementing anonymisation or pseudonymisation processes where required, and constructing and maintaining the data pipelines. They usually also work on ensuring that data environments meet agreed security standards, including access controls, logging, and auditability. Given the volume and velocity of MPD, experience with large-scale distributed systems and database optimisation is often required.
- **Data science and data analysis:** Data scientists and data analysts are responsible for transforming processed MPD into meaningful statistical outputs and decision-support tools. This includes the construction of indicators, calibration of models, validation against ground truth data, and development, deployment and documentation of methodologies used. The statistical rigor of their work will be critical to the final results, so they need to be familiar with inferential statistics, machine learning techniques (where appropriate) and reproducible research practices. They will also be needed to work closely with subject-matter experts to ensure that outputs are policy-relevant and methodologically defensible. They are also responsible for implementing clear version control, transparent codebases, and replicable workflows.
- **Survey expertise:** Where MPD is used to complement, calibrate or partially substitute traditional data sources, survey statisticians and sampling experts play a vital role. These professionals provide guidance on benchmarking MPD-derived indicators against other data such as field or phone surveys, assessing coverage bias, and designing hybrid methodologies that integrate conventional and non-traditional data sources. In many cases, they will work with analysts to develop weighting schemes, address biases inherent in mobile phone usage patterns, and assess representativeness relative to the target population. They can also support the development of val-

idation frameworks, advise on weighting and adjustment procedures, and help interpret discrepancies between sources (Cabrera and Rowe 2025). Their involvement is particularly important in official statistics contexts, where methodological transparency and compliance with statistical quality standards are required. Survey experts also help ensure that MPD outputs are aligned with established indicator definitions and international reporting frameworks where applicable.

- **Legal, data protection and compliance:** Given the sensitive nature of telecommunications data, legal and compliance expertise must be embedded in the initiative from its inception. Legal advisors are responsible for reviewing and drafting data sharing agreements, memoranda of understanding, and contractual provisions that define permissible use, retention periods, liability, and intellectual property arrangements. Data protection officers or privacy specialists should assess compliance with applicable data protection legislation and regulatory frameworks, including requirements related to anonymisation, purpose limitation, data minimisation, and user rights. They may also oversee data protection impact assessments and ensure that governance structures are formally documented. Close collaboration between legal, technical, and analytical teams is essential to ensure that privacy-preserving measures are technically feasible and legally robust.
- **Other roles:** In addition to the roles specified above, it can be useful, depending on the nature of the initiative, to include in the project team staff who are able to deliver communications activities, media liaison work, and monitoring and evaluation to gather information on how well the initiative is meeting its purpose. Communications specialists can help articulate the objectives, safeguards and benefits of the initiative to both internal and external audiences, which can be particularly important where public trust considerations arise. Monitoring and evaluation professionals can design frameworks to assess effectiveness, uptake, and impact, including the identification of performance indicators and feedback mechanisms. Finally, as with any project, administrative functions such as financial expertise will also be necessary and need to be identified and assigned to support the initiative. Budget oversight, procurement management, and resource tracking are essential to maintaining operational continuity and accountability.

1.4 Legal, Regulatory, and Ethical Context

MPD initiatives operate within a complex legal landscape that often spans multiple domains, including data protection, telecommunications regulation, cybersecurity, and national security law. Compliance with these frameworks is not optional and must be addressed as an integral part of planning, not as an afterthought.

Legal analysis should clarify the lawful basis for data processing, the roles and responsibilities of data controllers and processors, and the rights of data subjects. Because legal frameworks vary significantly across jurisdictions, including regional frameworks such as the African Union Malabo Convention, expert legal advice is essential ([African Union 2018](#)). Training materials and general guidance cannot substitute for context-specific legal interpretation. Ethical considerations extend beyond legal compliance.

Even where data use is lawful, it may raise concerns related to fairness, proportionality, discrimination, or surveillance. Ethical reflection should therefore accompany legal analysis, particularly for use cases involving vulnerable populations. ([GSMA 2016](#))

1.5 Stakeholders Ecosystem and Governance

MPD initiatives are inherently collaborative. There are a significant number of stakeholders who need to be involved, consulted and/or considered when planning and implementing an MPD initiative.

These include, for example:

- **National statistical offices (NSOs):** NSOs normally have a legal mandate and authority to produce a country's official statistics (in some contexts, additional agencies may also have such mandates). NSOs are also usually enabled through a legal basis to collect personal data (statistical authority) and are expected to abide by certain legal guarantees and obligations to protect it (statistical confidentiality). They will have teams with methodological expertise for producing statistical outputs and will also often have access to complementary data sources for validation and bias adjustment of CDR data. The NSO is often an end user of an MPD initiative's outputs, and may also be a conduit for distribution of such data to other parts of government.
- **MNOs:** MNOs are the data owners of CDR data. They control the underlying data and can provide access to it if they decide to do so, on the basis of certain considerations. MNOs tend to have good internal knowledge on the data source, information technology technical capacity and knowledge about the systems that produce and store the data. For ongoing, sustainable data pipelines, MNOs need to agree to provide regular access to new CDR data, including all associated data fields as well as regular updates to cell tower location data. When data gaps arise due to disruption to a data pipeline (e.g. power outages, infrastructure damage etc) they need to be able to help fill those data gaps. MNOs will seek to both protect subscriber privacy and their own commercial interests.

- **Telecommunications regulator:** The regulatory body responsible for oversight of MNOs and issuing licences to them already has regular contact with MNOs and should be consulted early in the process of establishing an MPD initiative in order to ensure that they have no objections to data processing. In some cases, the regulatory authority may itself have a desire to obtain insights from CDR analysis, as it may also have a development agenda. It may be that the regulator itself also has a mandate to collect data such as CDRs records. It can therefore play a number of different roles including: (a) facilitate a partnership in the initial phases; (b) processing CDRs to produce Information Society statistics; (c) stewarding data for other applications.
- **Data protection authorities:** Most countries now have a data protection commission, agency, bureau or similar body responsible for overseeing and regulating data protection. Engaging with this regulatory body early in the design of an MPD initiative is sometimes required by law but in any case is advisable, as it can help to ensure that public interest objectives are balanced with privacy and rights protections, thereby strengthening public trust. These authorities can also give advice and guidance on data protection measures and whether they are proportional to the risks.

In addition to the above, whose involvement is normally essential for a sustainable initiative, the following stakeholders should also be considered:

- **Users:** Any end-user, such as a Ministry, Department, Agency or other actor within government or beyond, should be involved at an early stage in order to effectively understand their needs, and specify their requirements. It's also possible that their involvement can provide support to the planned activities, for instance through mobilising political support for the initiative or helping to raise funds to proceed;
- **Research community:** The academic community can provide useful inputs to MPD initiatives by supporting development of methods, providing academic rigour and mobilising resources for research;
- **Technical service providers:** It can be beneficial to consult and/or engage the services of experts in the field who have international exposure, can share experiences and best practices, and provide technical assistance and solutions or products;

Each of the above stakeholders has their own particular mandate, incentive structures, and concerns they may have in relation to using MPD. Addressing these will be essential to ensure ongoing cooperation within the data ecosystem. So too will creating the necessary framework within which the relevant stakeholders will work together. This requires that effective governance be put in place, with clearly defined roles, decision-making processes, and accountability mechanisms that reflect the diversity of stakeholder interests. For more on good data governance practices, see Chapter 6.

i Box 2: The Maturity Assessment Framework: a Tool to Assess Readiness and Progress

As part of the World Bank's Global Data Facility's MPD for Policy programme, a [Maturity Assessment Framework](#) was developed which provides a structured yet flexible way to assess readiness, progress, impact, and sustainability of MPD initiatives for official statistics. Built around four maturity stages and three assessment areas (feasibility, impactfulness, and sustainability) the framework helps to identify strengths, gaps, and support needs across the legal, technical, organisational, ethical, and financial dimensions briefly described in this chapter. The framework is accompanied by digital and printable self-assessment tools, and is intended to support individual reflection, stakeholder dialogue, and informed planning, while allowing adaptation to local contexts. Readers are encouraged to download a copy of the tool and use it to assess the current state of play in their context as well as envision the future state desired (linking this to a Theory of Change exercise as described in Box 1).

1.6 Risk management and mitigation measures

Planners of MPD initiatives need to consider and explicitly address a range of risks which are often interconnected and can quickly undermine an initiative if not proactively managed. We focus here on the data-related risks associated with processing CDRs but planners should also consider and plan to mitigate risks such as ethical risks (including bias, exclusion, or misuse); reputational risks such as those arising from security breaches, public misunderstanding or mistrust in use of MPD for by non-MNO bodies; and legal and regulatory risks such as those associated with non-compliance with licensing conditions or breaking local laws.

1.6.1 Data-related risks

There are a number of risks associated with MPD initiatives related to subscriber privacy.

1) Privacy Risks

This includes the risk of invading personal data privacy through unauthorised access to sensitive personal information, especially individual-level MPD or other personal information held by operators as well as the risk of MPD being used for the surveillance or profiling, including social, financial or behavioural profiling, of subscribers.

2) Security Risks

One way in which unauthorised access to data can occur is through data breaches. There are a number of different types of data breach that need to be considered when implementing an MPD initiative. Data breaches may occur as a result of malicious actors exploiting vulnerabilities in the cybersecurity of an MPD pipeline (“hacking”) in order to gain unauthorised access to data. However, unauthorised access may also result from the improper storage or sharing of data. This can include insecure storage or transmission of data or the intentional or unintentional sharing of sensitive data with unauthorised parties. Data breaches can also involve data availability or integrity breaches, where data is intentionally or unintentionally destroyed, corrupted or altered in a way that is inconsistent with the established processes. This might include the accidental or malicious alteration or deletion of records from a database or the destruction or malfunctioning of the server storing the data.

3) Ethical Risks

This refers to the risk of MPD being used for unethical purposes, either by internal or external actors which may be intentional, malicious or accidental. MPD may be abused to target certain populations with the intention of doing harm to or inequitably benefit given groups. It’s also a significant risk that, without proper methods to address the inherent biases in MPD, parts of a population will also be over- or under-represented in the data produced, resulting in biased decision-making which could - for instance - exclude specific groups from benefits (e.g. access to resources) or expose them to harm (e.g. targeting of internally displaced persons).

1.6.2 Risk mitigation measures

Risk mitigation can be done through a combination of technical, organisational, and governance measures. These include clear staff responsibilities, regular training in data protection and cybersecurity, and application of privacy-enhancing techniques. Engaging oversight bodies, ethics committees, and civil society can further strengthen legitimacy and resilience.

It is good practice for MPD initiatives to develop and implement a robust data governance framework which incorporates strong measures to assess privacy, security and ethical risks and identify appropriate mitigations. The process of assessing and mitigating risks should be transparent and ideally also inclusive of the different stakeholders in an MPD initiative.

Some of the other mitigation strategies that can be adopted by an MPD initiative include:

- Effective use of robust data encryption tools
- Strict, granular access controls, monitoring and auditing
- Regular security updates and security audits (this is important, because risks evolve over time with changes in technology, personnel, and legal frameworks)
- Full compliance with all relevant data protection regulations.

1.7 Designing for Long-Term Sustainability

Sustainability should be considered from the earliest planning stages. Short-term pilot projects can generate valuable learning, but lasting impact requires long-term data access arrangements, scalable infrastructure, and institutionalised processes.

Key elements of sustainable design include:

- Clearly defined and enduring objectives
- Robust data governance frameworks (see also Chapter 6)
- Automated and resilient data pipelines (see also Chapter 4)
- Continuous investment in human capacity
- Ongoing stakeholder engagement and communication (see also Chapter 7)
- Integrated monitoring, evaluation, and learning systems

Financial sustainability also matters. While initial setup costs may be high, planners must account for ongoing operational expenses, including infrastructure maintenance, staff retention, training, and stakeholder engagement. Funding strategies may need to combine internal budgets, in-kind contributions, donor support, and the use of open-source tools.

1.8 Conclusion

Planning an MPD initiative is a complex, multidisciplinary undertaking. Technical feasibility, legal compliance, ethical responsibility, institutional coordination, and long-term sustainability are all equally important. By approaching planning as a structured, purpose-driven process and by investing in both people and systems, organisations can responsibly harness MPD to generate meaningful public value.

2 Chapter 2: Policy Applications for MPD

In recent years MPD has emerged as a transformative resource for governments, NSOs, and development partners who want timely, granular, and cost-effective forms of evidence to inform their policy-making. This chapter addresses the use of MPD for policy-making with an emphasis on policy applications in which MPD has already been demonstrated to add value (Deville et al. 2014; Lu et al. 2012; Blumenstock et al. 2015). This chapter discusses *why* MPD is valuable, *how* it can be applied across key policy domains, and some *considerations* that must be addressed when designing an MPD initiative to achieve particular policy-informing objectives.

2.1 Why and How is MPD Useful?

As discussed in Chapter 1, MPD provides a continuous, passively collected record of population presence and mobility (Louail et al. 2014). Unlike traditional surveys or censuses, which are costly, infrequent, and static, MPD enables:

- **High temporal resolution:** daily or even near-real-time indicators
- **Fine spatial granularity:** insights at neighbourhood, district, or corridor level
- **Cost efficiency:** lower marginal costs once partnerships and pipelines are established
- **Resilience:** data collection continues during crises such as pandemics, disasters, or conflicts

These characteristics make MPD particularly valuable in policy areas where *where people are* and *how they move* directly affects outcomes (Gonzalez et al. 2008; Song et al. 2010; Deville et al. 2014).

The UN Committee of Experts on Big Data and Data Science for Official Statistics has an MPD task team which has published guidance on using this data source for different specific use cases (United Nations Committee of Experts on Big Data and Data Science for Official Statistics n.d.). The task team has identified six domains where MPD is especially relevant and being actively used in different contexts:

1. Measuring the Information Society
2. Dynamic population mapping and migration statistics
3. Disaster management, displacement and public health crises
4. Tourism statistics
5. Transport statistics

For each of the above, detailed guides have been produced which provide in-depth information regarding how to approach using MPD for these purposes. For a comprehensive description of each, please refer to the specific topic's Guide (see links provided in Appendix 1). The next section gives overview summaries for each with some illustrative examples. It is followed by a short description of another potential application for MPD, estimating the socio-economic variables such as distribution of wealth and poverty, which is being actively explored.

2.2 Information Society Statistics

2.2.1 Information Society and Digital Inclusion

Access to digital technologies is now a core determinant of social and economic inclusion. Yet traditional household surveys often struggle to provide timely and geographically detailed information on Internet use and mobile connectivity. MPD can support SDG monitoring by producing indicators related to: ([UN-CEBD Task Team on Mobile Phone Data n.d.c](#))

- **Internet usage** (SDG 17.8.1): proportion of individuals using the internet
- **Network coverage** (SDG 9.C.1): proportion of the population covered by mobile networks

Practical Guidance When using MPD for digital inclusion indicators:

- Ensure the operator's subscriber base is sufficiently representative of the population
- Adjust for **technology type** (2G, 3G, 4G, 5G), as not all devices support internet use
- Validate MPD-derived indicators against household surveys where possible

Box 3: Brief Case Studies of Information Society applications ([International Telecommunication Union n.d.a](#))

Rio de Janeiro, Brazil: MPD-based estimates of internet use differed by only 1% from household survey results at metropolitan level, demonstrating that MPD can reliably approximate traditional indicators while offering much finer geographic detail. **Bali, Indonesia (2020):** MPD slightly overestimated internet use compared to socio-economic surveys, highlighting the importance of accounting for older technologies such as 2G phones that cannot access the internet.

2.2.2 Network Coverage and Infrastructure Planning

MPD allows policymakers to distinguish between nominal coverage and *effective* access to high-speed connectivity. Mapping 2G versus 3G coverage, for example, reveals where populations may technically be covered but lack access to broadband-quality services. Such insights directly inform:

- Infrastructure investment prioritisation
- Universal service obligations
- Digital equity strategies

2.3 Population Statistics and Mobility Analysis

Accurate population data is critical for governments to plan services, allocate resources, and respond effectively to emergencies. In addition, accurate population data forms the basis of most other statistics, whether providing a reference base for the calculation of statistics or setting the frames for surveys. MPD can be used in different ways to understand and use population movement data.

2.3.1 Dynamic Population Mapping and Migration

Dynamic Population Mapping uses MPD to estimate where people are located at different times of day or year, capturing the *de facto* population rather than the static, *de jure* population recorded in censuses. This approach captures the actual presence and movement of populations over time. (UN-CEBD Task Team on Mobile Phone Data n.d.b; Deville et al. 2014; Ricciato et al. 2020)

Dynamic population mapping is useful for several use cases, including:

- Service provision (health, policing, transport)
- Infrastructure planning
- Emergency preparedness
- Event and seasonal population management
- Creation of dynamic sample frames for surveys
- Census preparation and implementation

MPD is not designed to entirely replace conducting a census. Rather, it can be used to strengthen such data collection activities by, among other things: (a) Assisting in production of sample frames or enumeration areas; (b) Identifying populations that have been, or are at risk of being, undercounted; and (c) Providing interim updates between census rounds. When using MPD for such use cases, it is critical that planners explicitly address **bias risks**, given that mobile phone ownership is lower among children, the elderly, women in some contexts, and poorer households (Wesolowski et al. 2013; Cabrera and Rowe 2025).

Some of the key design principles for using MPD for dynamic population mapping include:

- Use sufficiently long historical windows to distinguish residents from visitors
- Segment populations (residents, commuters, tourists, transit users)
- Validate estimates against administrative or survey benchmarks

Estimating the movements of populations in a country can also be taken a step further, and used to produce official migration statistics. The UN-CEBD Task Team on MPD has produced methodological guidance on how to do so. ([Rowe et al. 2022](#))

i Box 4: Dynamic population data and internal migration

Estonia (2012 to 2015): From 2012 to 2015, the Estonian Police and Border Guard needed accurate and up-to-date population statistics for all 237 municipalities. Their goal was to better plan the distribution of patrol units across the country. To do this, they needed more than just total population counts. They needed to understand seasonal and weekly fluctuations, and distinguish between permanent residents, domestic visitors such as workers and students, foreign visitors, and even people in transit. They turned to MPD, or MPD, to estimate the de facto population. By looking at longer-term patterns, MPD also allowed them to differentiate residents from visitors and people in transit. The rescue services of Estonia repeated the exercise with data for 2018 as well as 2021-2022. Madrid (2019, 2020): The project aimed to analyse migratory movements in the Autonomous Community of Madrid during the COVID-19 and post-pandemic periods using MPD. It found evidence of shifts in migration patterns, including increased inflows to rural and outer suburban areas and outflows from core urban areas, across different population groups.

2.3.2 Disaster Management

Traditional statistics struggle to measure displacement and return dynamics. MPD can enable rapid estimation of displacement, return, and recovery when digital trace data are carefully adjusted, validated, and triangulated with other sources ([UN-CEBD Task Team on Mobile Phone Data n.d.a](#); [Lu et al. 2012](#); [Rowe 2022](#); [Iraddock et al. 2025](#); [Pietrostefani et al. 2025](#)). MPD can enable rapid estimation of:

- The scale of displacement
- Destinations and duration of movements
- Return and recovery patterns

As a result, MPD can be a very useful source of information for responding to disasters. How populations move after large and small-scale events, from earthquakes and hurricanes that destroy vast areas, to small-scale localised conflicts such as gang violence, the data can support an understanding of how populations who have been affected are responding in terms of movement out of specific geographies, where they are going to, how many are displaced, and for how long they are displaced. This data is extremely useful to organisations

wanting to respond to disasters with needs assessments; emergency response interventions such as shelter, food, and health services; and cash transfers to affected populations. For example, during the gang violence in Haiti, information provided through the Haiti Mobility Data Platform since 2023 has been used by UN agencies and the humanitarian NGOs operating in the country to identify and where to send assessment teams, what parts of the country to prioritise for funding and assistance. Another example comes from Ghana, where Ghana's National Disaster Management Organisation, NADMO, received reports on displaced populations after an initial response to flooding in the Lower Volta region in 2023. The assessment identified a population of affected people that had previously not been identified and to which they could send an assessment team to clarify needs and provide assistance.

MPD can also be used to assess the impact of early warning, evacuation alerts and other government measures (e.g. lockdowns during COVID) (Li et al. 2021; Rowe et al. 2023). For example, during the 2024 wildfires in Valparaíso, Chile, MPD from approximately 580,000 devices was analysed to evaluate the effectiveness of SMS messages warning the population about the fires and informing those in risk areas to evacuate. The data enabled high-frequency observation of evacuation behaviour at operational timescales.

Finally, MPD can be used to understand connectivity access as well as mobility disruptions during an emergency. For example, during the 2025 Spain–Portugal blackout, MPD revealed how many people were away from home and how far they were displaced at the time of the outage.

i Box 5: Brief Case Studies of MPD for disaster management

Madrid during COVID-19 (2020): MPD showed a 10% drop in population during lockdowns and identified destination regions, helping assess societal impacts of mobility restrictions. Ghana during COVID-19 (2020): Data from Vodafone Ghana was used to support government decision making by the Presidential Task Force around the effectiveness of COVID-19 lockdowns and what non-pharmaceutical interventions were working in the country, informing subsequent policies on movement restrictions (Li et al. 2021). Bangladesh (climate risk): MPD quantified migration linked to cyclones and environmental stress, informing adaptation and resilience strategies. Haiti earthquakes (2010, 2021): MPD provided rapid, reliable estimates of displacement that closely matched later survey results, supporting both emergency response and long-term recovery planning. Chile (2024): Researchers analysed MPD to assess how effective SMS evacuation messages had been. Haiti gang violence (2021-2025): MPD was used to demonstrate how the population affected by gang violence was relocating from areas affected, particularly in the capital, but moving to areas that are more prone to flooding during hurricane season. It was also used to study the effect on movements following announcement and deployments of peace and security missions to Haiti.

2.3.3 Public Health and Epidemiology

MPD offers public health practitioners a powerful input for epidemiological surveillance and outbreak response. Because infectious disease spreads through human movement, it can provide a near-continuous, large-scale record of population mobility that can be used to model spatial transmission risk, identify hotspots, and predict where outbreaks are likely to emerge next. This has been demonstrated across multiple disease contexts: CDR-derived mobility metrics outperformed conventional gravity models in predicting the spread of cholera in Haiti in 2010 and revealed the role of mass gatherings as a transmission driver during the 2005 Senegal epidemic ([Bengtsson et al. 2015](#); [Wesolowski et al. 2012](#); [Tizzoni et al. 2014](#)). Similar approaches have been applied to rubella in Kenya, where seasonal mobility patterns tied to school terms and holidays outperformed other variables in explaining transmission peaks, with direct implications for vaccination timing.

Beyond spread modelling, MPD enables real-time evaluation of non-pharmaceutical interventions such as travel restrictions and lockdowns, providing evidence on behavioural compliance that cannot be obtained from any other source. During the Ebola response in Sierra Leone and COVID-19 responses across multiple countries, CDR and other digital trace analyses showed measurable reductions in mobility following restrictions (and their reversal once measures were lifted), often within just a few days of implementation ([Rowe et al. 2023](#); [Cabrera et al. 2025](#)).

MPD-derived population statistics can also be used to support health monitoring and metrics by producing dynamic population denominators which reflect actual population density numbers, rather than static census counts ([Deville et al. 2014](#); [Ricciato et al. 2020](#)). In Ghana, WFP supported analytical work by the Data for Good Partnership working with Ghana Health Service and Ministry of Health colleagues to combine CDR-based mobility estimates with disease case counts and generate per-capita indicators for resource allocation and outbreak preparedness which took dynamic population movements into account.

By integrating MPD with environmental and health datasets, authorities can generate dynamic exposure indicators that reflect actual population movements. Some applications include:

- Disease modelling
- Environmental surveillance of diseases (e.g. polio)
- Emergency health service planning
- Pollution and heat exposure

i Box 6: Case study of MPD for Public Health and Epidemiology

Haiti cholera outbreaks (2010, 2022): After the 2010 earthquake in Haiti, MPD was used to demonstrate how MPD-derived indicators of movement can be used to predict the spread of cholera; and in 2022 further analysis was done of how outbreaks would cause infectious pressure in specific parts of the country based on population movements. MePreCISa Project (2024, 2025): The MePreCISa Project, led by the Barcelona Supercomputing Center (BSC), developed an open cloud platform designed to support the management of health crises in complex scenarios by integrating MPD with health and environmental information to model pollution exposure and disease spread. It addresses key use cases such as air quality and health, social contact and epidemic transmission, and mobility and wastewater in the Autonomous Community of Catalonia. India Tuberculosis Spread (GSMA/Airtel): A GSMA study used anonymised Airtel CDRs to model and track the spread of tuberculosis (TB) in India, helping to identify high-risk areas and inform targeted public health interventions.

2.3.4 Tourism Statistics

MPD can help to overcome some of the common limitations of tourism surveys by capturing cross-border and short-duration movements. ([UN-CEBD Task Team on Mobile Phone Data n.d.d](#); [Ahas et al. 2008](#)) The specific ways in which MPD can be used for tourism statistics and data include:

Measuring inbound and cross-border visitor arrivals. The most widespread use of MPD in tourism is to measure the number of people crossing international borders. This is particularly valuable where traditional border controls have been removed or weakened (see Estonia case study).

Measuring outbound travel. MPD can equally capture the movement of residents travelling abroad. Estonia uses MPD to produce both inbound and outbound border crossing statistics, which feed directly into the Travel Services component of the Balance of Payments; a use case that illustrates how MPD contributes not just to tourism policy but also to macroeconomic accounting.

Domestic tourism measurement. Beyond international travel, MPD has also been used to measure trips taken by residents within their own country. Indonesia, for example, combines MPD with digital surveys to track domestic tourist trip volumes and purpose of travel.

Visitor profiling and disaggregation. Granular analysis of the profiles of tourists can also be helpful, and MPD can allow analysis of inflows and outflows of subscribers by, for instance, country or region of residence, type of visit (same-day visitor, overnight tourist, transit visitor), and geographic destination within the host country.

Length of stay estimation. By tracking the duration of a mobile device's presence within a defined geographic area, MPD can estimate how long visitors remain in a destination, at national or sub-national level.

Event and venue visitor measurement. MPD has been applied to measure visitor flows at specific events and locations. For instance, during the 2018 Asian Games in Indonesia, MPD was used to count attendees at venues in Jakarta and Palembang, identify visitor origins, track mobility between venues and cities, and estimate length of stay. This information could not be reliably produced through ticket sales, immigration records, or surveys alone. This visitor data was then used as an input into a Computable General Equilibrium (CGE) model for economic impact analysis.

Supporting official statistics and policy planning. Across countries, a shared motivation for adopting MPD is to improve the timeliness, coverage and cost-efficiency of official statistics. Benefits can include timeliness, consistency, completeness of coverage and cost-effectiveness. MPD-derived tourism statistics can also feed upstream into national development planning, balance of payments, and sustainable tourism policy, including progress monitoring against SDG targets.

i Box 7: MPD for tourism statistics

Indonesia (2016, 2018, 2019): MPD was used for inbound tourism, domestic tourism, outbound tourism, and event-impact analysis. It reduced work burdens, increased granularity from province to city/municipality level, and reduced the budget (by more than half). City/Municipality Local Tourism Authority used the data for development planning, investment and promoting local tourism spots. MPD also provided data on country of destination and increased the coverage especially at the areas where there is no immigration checkpoint (Outbound Tourism 2019, official statistics). The local and national authority use it for policy on increasing domestic tourism. In 2018, MPD is also used in study for the Impact of Asian Games 2018 on Indonesia (National/Regional) Economy. The results of the study is used to develop tourism nationally and regionally. Estonia (2008 onwards): Estonia's Bank of Estonia (Eesti Pank) began using MPD to compile border crossing statistics in 2008, prompted by Estonia's entry into the Schengen area which eliminated traditional border controls, and budget constraints that shifted responsibility for travel statistics away from Statistics Estonia. A methodology was developed to estimate the number of inbound and outbound travelers, trip durations, and visitor categories, feeding these figures into the Balance of Payments Current Account. Data is collected monthly from MNOs under a legal mandate requiring companies to supply data for official statistics, processed into anonymised aggregate outputs, and validated against payment card data, airport statistics, and accommodation surveys. Results have been published quarterly as "International Travel Statistics" since 2012.

2.3.5 Transport and Mobility Planning

Designing effective transport policies requires a comprehensive understanding of how people and goods move. Information on travel demand, such as origin–destination patterns, trip frequency, distance, timing, and modal choices; is essential for infrastructure planning, service design, and policy evaluation. It also plays a critical role in monitoring progress towards sustainability objectives, including those related to climate change mitigation, air quality, and equitable access.

In the transport arena, MPD has been applied to a wide range of policy-relevant use cases. These include:

- the development and monitoring of national and urban mobility plans,
- the evaluation of transport infrastructure investments,
- the optimization of public transport services, and
- the assessment of policy measures such as congestion management or low-emission zones, among others.

Case studies, particularly from Latin America and Europe, have demonstrated the strong value of MPD, highlighting improvements in the quality, spatial coverage, and temporal continuity of mobility statistics, as well as the ability to analyse mobility patterns over multiple days and time periods. (Caceres et al. 2007; Calabrese et al. 2011; Alexander et al. 2015; Toole et al. 2015)

MPD offers several advantages compared to traditional data collection methods. Its large sample size, often covering a significant share of the population, enhances representativeness and enables detailed spatial analysis, particularly for the estimation of origin–destination matrices (Caceres et al. 2007; Calabrese et al. 2011; Toole et al. 2015). In addition, MPD provides continuous observations over time, allowing for more frequent updates and enabling the monitoring of mobility dynamics, including responses to transport policy interventions such as congestion pricing schemes, public transport reforms, or mobility restrictions. When combined with other data sources, such as surveys or smart card data, MPD can also support more comprehensive analyses, including modal segmentation.

i Box 8: MPD for transport statistics

Ministry of Transport and Sustainable Mobility (MITMS), Spain (2017-2025). Within its responsibility for planning and managing transport infrastructure and services in Spain, the MITMS conducts so-called “Demand Prospective Studies” to assess the impact of new infrastructures, services, and regulations on the transport system. One of these focuses on analysing passenger mobility at the national level. Traditionally, this was carried out through the Movilia surveys, which required significant economic, technical, and human resources. To leverage emerging mobility data sources, in 2017 the MITMS launched the “Analysis of Interurban Mobility in Spain using Big Data Technologies,” a pioneering project based on MPD to study mobility patterns nation-

wide. This initiative demonstrated that MPD can provide high-quality mobility insights at a lower cost and with greater timeliness than traditional surveys. The project was extended in 2020–2021 to enable daily mobility monitoring and support decision-making during the COVID-19 crisis, and further consolidated in 2022–2025 with a continuous mobility analysis framework, integrating MPD with additional data sources, such as public transport ticketing data, to monitor the evolution of travel demand behaviour at national level. Metropolitan areas in Latin America (2021–2026). Since September 2021, the World Bank has used MPD to generate origin–destination matrices for several metropolitan areas in Latin America, with the overarching goal of providing strategic guidance and technical recommendations for the development of a regional transport planning platform in the LAC region. These projects aim to deliver high-quality, up-to-date passenger travel demand information based on new big data sources. Initially implemented in Bogotá (Colombia), Buenos Aires (Argentina), and Medellín (Colombia), the initiative was later extended to Asunción (Paraguay) in 2023 and to Rio de Janeiro, Belo Horizonte, and Florianópolis (Brazil) in 2025. Department for Transport UK: The UK Department for Transport (DfT) used MPD to understand national and regional travel patterns, particularly in response to major transport disruptions or for informing long-term strategic road network planning. This work, often referenced in the UN Handbook series, focused on deriving Origin-Destination (OD) matrices to gain timely insights into commuter and non-commuter movements, complementing traditional travel surveys with high-frequency, large-scale data.

2.4 Socio-Economic Applications: Poverty Mapping

Traditional poverty data can often become quickly outdated, particularly in low- and middle-income countries and contexts with highly dynamic populations. In combination with traditional sources such as census and survey data, in periods between their data collection, and sometimes in combination with other forms of data such as geospatial datasets, MPD can be a useful tool for generating updated and spatially refined estimates of socio-economic variables such as wealth or poverty ([Eagle et al. 2010](#); [Blumenstock et al. 2015](#)).

Work in this area is still relatively nascent with exploratory analysis and in-depth research being undertaken by a number of researchers and institutions. Some initial findings and examples of application to real-world scenarios are described below.

2.4.1 Bangladesh poverty mapping study

In Bangladesh, researchers have demonstrated how anonymised mobile phone metadata such as CDRs capturing patterns of mobility, airtime purchases, and social connectivity can be leveraged to estimate poverty at a much finer spatial resolution than traditional household surveys allow. By statistically linking these behavioural proxies to benchmark survey-based welfare measures, the studies produced high-resolution poverty maps capable of identifying significant regional disparities, including pockets of deprivation that are often masked in national or district-level averages.

These poverty maps proved particularly valuable for policy analysis and planning, as they enabled near-real time updates and granular geographic targeting at relatively low marginal cost. In a context where surveys are expensive and infrequent, mobile phone based approaches complemented official statistics, offering decision-makers an evidence base to prioritise lagging regions, allocate resources more efficiently, and monitor spatial inequality over time. ([Blumenstock et al. 2015](#); [Steele et al. 2017](#))

2.4.2 Togo's use of MPD for targeting social protection payments

In Togo, MPD played a more operational role during the COVID-19 pandemic, when rapid identification of vulnerable populations was essential. Anonymised mobile phone usage indicators were combined with geospatial data such as night-time lights and population density to infer economic vulnerability in areas where up-to-date poverty data were unavailable. This approach allowed authorities to estimate need dynamically as the crisis evolved.

The resulting analytics underpinned a digitally delivered social protection programme that targeted informal workers and low-income households for emergency cash transfers. By using mobile phone-based proxies rather than relying solely on existing registries, Togo was able to expand coverage quickly and transparently, demonstrating how mobile data, when integrated with geospatial information, can support timely and adaptive social protection in crisis settings. ([Aiken et al. 2022](#); [World Bank 2021](#))

2.4.3 Testing the utility of MPD for poverty prediction in Ghana

In 2024/25, a study was undertaken to explore how MPD can be used in model-based estimates of multi-dimensional poverty measures for Ghana. The analysis tested various data types (CDR-derived variables, geospatial data such as building footprints, and other forms of data from MNOs such as top-up data) to see which would be good predictors of multi-dimensional poverty indices (MPI) as estimated through census and survey instruments. The study found that covariates based on building footprints were consistently the strongest predictor for Ghana MPI; models built using a combination of CDR and geospatial covariates consistently provide marginally better performance than models using geospatial covariates alone; and models built using CDR covariates alone have consistently lower performance. Further areas of study have been identified.

2.5 Conclusion

MPD offers a powerful complement to traditional statistical systems. When carefully planned, validated, and combined with other data sources, MPD initiatives can significantly enhance governments' ability to understand populations, respond to crises, and design evidence-based policies. This manual provides a foundation for practitioners to move from isolated pilots toward sustainable, policy-relevant MPD programmes.

3 Chapter 3: Arranging Partnerships and Data Access

Planning an MPD initiative requires far more than technical capability or analytical ambition. At its core, such an initiative is an exercise in partnership-building, trust management, and institutional alignment. MPD is held by private-sector MNOs, is legally sensitive, and is embedded in complex commercial, regulatory, and public accountability environments. As a result, the success or failure of an initiative is often determined less by methodological sophistication than by how well partnerships are arranged and how data access is negotiated, governed, and sustained. This chapter is intended to guide practitioners, particularly those in NSOs or public-sector policy institutions, through the practical realities of arranging partnerships and securing appropriate data access models. It summarises the key steps and signposts additional useful considerations as well as resources. ([Global Partnership for Sustainable Development Data and Positium 2025](#))

3.1 Understanding the Central Role of MNOs

Any MPD initiative must begin with a clear understanding of the role of the MNO. Operators are not simply data providers; they are keystone stakeholders whose engagement determines whether a project can move beyond concept to implementation. Operators are the primary data holders. Through the operation of their networks, they collect and store CDRs or signaling data for their own operational and business purposes ([United Nations Statistics Division 2019](#); [GSMA 2019](#)). This gives them several forms of indispensable capacity: deep internal knowledge of how the data are generated, technical infrastructure capable of handling very large data volumes, and operational control over the pipelines that deliver data on an ongoing basis. For a statistical or policy initiative to function reliably, the operator must be able to provide regular access to new data, maintain accurate and up-to-date cell tower metadata, and intervene quickly when data pipelines fail or gaps emerge. The extent of the operator's role may vary depending on country context, regulatory frameworks, and institutional maturity, but their centrality does not. Effective planning therefore requires practitioners to explicitly consider the operator's perspective from the outset, rather than treating data access as a purely administrative or legal hurdle.

3.2 Why Operators May Hesitate to Share Data

Before approaching an operator, it is essential to understand why data sharing is often perceived as risky or unattractive from the operator's point of view. These concerns typically fall into three broad categories: compliance risks, capacity constraints, and business considerations. Recognising these concerns is not a concession; it is a prerequisite for designing an engagement strategy that is realistic and credible (GSMA 2019).

From a compliance perspective, operators are acutely sensitive to legal and reputational risk. Telecommunications data are subject to sector-specific regulation as well as general data protection and privacy laws. Even when a proposed use is lawful, operators may fear public backlash or civil society criticism if subscriber data are perceived to be misused (Jansen et al. 2021; Montjoye et al. 2018). This risk is amplified in environments where public trust in data governance is fragile or where the legal framework for public-private data sharing is ambiguous.

Capacity-related concerns are also common. While operators operate sophisticated IT systems, these systems are designed primarily for network management and commercial analytics, not for producing official statistics. An operator may lack the infrastructure to extract, store, or process the specific data fields required for statistical purposes, or may not have staff with the time or expertise to actively participate in a complex, long-term project.

Finally, business considerations can present significant barriers. Participation may require investments in new hardware, software, or human resources. In addition, many operators are developing their own data monetisation strategies. From this perspective, sharing data for free with public institutions may appear to undermine potential revenue streams or create competition with existing commercial products.

3.3 Managing and Mitigating MNOs' Risks

Although these risks are real, they are not insurmountable. A central lesson from practice is that most MNOs' concerns can be mitigated through careful project design and clear institutional arrangements.

Compliance risks can be addressed through robust legal agreements and data governance frameworks. Where the legal environment is unclear, contracts can specify roles, responsibilities, and safeguards in detail, thereby reducing uncertainty. In some contexts, risk is further reduced when the statistical authority has a clear legal mandate and can make participation mandatory. In such cases, the legal responsibility for compliance shifts away from the operator and toward the public authority, which many operators view as a significant advantage.

Capacity constraints can be managed by embedding training and support into the project design. This may include formal training courses, on-site technical assistance, or the use of third-party technical service providers to perform specialised tasks. In effect, capacity that does not exist within the operator can be supplemented or replaced through external expertise.

Business concerns require the most careful handling. While costs and effort cannot be eliminated, they can be offset by clearly articulated benefits. Crucially, these benefits do not need to be financial. In many successful projects, one or two well-defined non-monetary incentives have been sufficient to secure sustained cooperation.

3.4 Articulating Clear Incentives for MNO Participation

Experience shows that operators are far more willing to engage when they can clearly see how participation aligns with their interests. These incentives tend to fall into two broad groups: benefits arising from collaboration with a statistical authority, and longer-term capacity and innovation benefits.

Key collaboration-related benefits include:

- **Regulatory de-risking:** When a statistical authority assumes legal responsibility for data use, particularly in mandatory data-sharing arrangements, the operator's exposure to regulatory and compliance risk is reduced.
- **Public recognition and corporate social responsibility:** Participation in projects with visible public value allows operators to demonstrate positive societal impact. This can be reflected in corporate social responsibility reporting and public communications.
- **Data exchange opportunities:** Statistical offices often hold high-quality demographic and socio-economic data from censuses or surveys. When appropriate, sharing aggregated or derived insights with operators can help them improve marketing strategies, business planning, and infrastructure investment decisions.

In addition to these immediate benefits, MPD initiatives can support longer-term capacity building within operators. High-quality statistical projects often involve international experts, rigorous methodologies, and strict quality standards. By participating, operators can:

- Develop advanced analytical capacity that can later be applied to commercial products.
- Build infrastructure that supports both public-interest and private-sector analytics.
- Improve the statistical robustness and credibility of their own data products by incorporating principles such as representativeness and quality assurance.

For practitioners, the key task is not to offer all possible benefits, but to identify which incentives are most compelling for a specific operator and to focus negotiations accordingly.

3.5 Defining Boundaries Between Public and Private Data Use

A frequent source of tension in negotiations is concern about market competition. Statistical offices produce public goods- official statistics intended for broad public use. Meanwhile, operators are private sector operators whose incentive structures mean they will often seek to monetise data and detailed insights for specific clients. Successful initiatives explicitly address this issue by agreeing on a clear division of these markets.

In practice, this means that the statistical authority commits to producing high-level, validated indicators that serve the public interest and are disseminated openly, typically at relatively coarse levels of disaggregation. At the same time, the operator retains the right to develop and sell more granular, client-specific products that do not require the same level of statistical validation.

For example, an NSO or Ministry of Transport might publish monthly commuting patterns between cities, while the operator offers daily or hourly movement data between neighbourhoods for commercial customers. By making these boundaries explicit, both parties can operate in the same data ecosystem without undermining each other's objectives.

3.6 Understanding Data, Technical, and Legal Roles

Planning data access requires clarity about both technical and legal roles. MPD projects involve multiple datasets, ranging from highly sensitive raw databases to publishable aggregate indicators. As data move along this processing chain, their sensitivity decreases, but governance requirements remain strict. From a technical perspective, responsibilities typically include:

- **Data collection**, usually performed by the operator through its network operations.
- **Data processing design**, often led by the statistical office, sometimes in collaboration with research institutions or technical service providers, and focused on methodology and quality frameworks.
- **Data processing**, which may be carried out by the operator, a third-party provider, or in some cases the statistical office itself.
- **Data evaluation, validation, and dissemination**, generally the responsibility of the statistical office or a relevant government ministry.

Alongside these technical roles are legal roles defined by data protection frameworks. The data controller determines the purposes and means of processing personal data, while data processors act on the controller's behalf under contract. Determining who can legally assume these roles, and under what conditions, is central to selecting an appropriate data access model.

3.7 Using Maturity Assessment to Select a Data Access Model

There is no single “best” data access model. Rather, the appropriate model depends on institutional maturity, assessed across three key dimensions. First, legal authority: does the receiving party have a clear mandate to collect and control private-sector data, and can it compel data sharing if necessary? Second, technical capacity: does it have the infrastructure and expertise to securely store and process massive volumes of raw data, potentially from multiple operators? Third, public trust: does the institution enjoy sufficient credibility and political support to act as a steward of highly sensitive data?

Depending on the answers to these questions, the statistical office may operate as:

- A **data user**, receiving final aggregate indicators produced by an MNO or technical partner.
-
- A **data controller without being a data processor**, specifying the needs and data processing mechanisms for statistical production that are performed by the MNO or technical provider. The NSO performs the validation, dissemination and evaluation tasks.
-
- A **data controller and a data processor**, performing the above tasks, while collecting and processing raw data within its own secure environment or through authorised processors.

		Data Protection Roles		
		Data Controller	Data Processor	Data Owner
Business Process	Specify needs Design Build Collect Process Analyze Disseminate Evaluate Data User			

Each model has been used successfully in different country contexts, and it’s also been the case that initiatives evolve over time as legal frameworks, capacity, and trust develop.

3.8 Formalising Partnerships Through Agreements

Regardless of the chosen model, partnerships must eventually be formalised. Early stages may rely on non-disclosure agreements or memoranda of understanding, but sustained data access typically requires detailed data-sharing agreements.

These agreements should clearly address issues such as anonymisation, data handling procedures, permitted uses, access controls, retention and deletion policies, and compliance with applicable data protection laws (GSMA 2019; Montjoye et al. 2018). Well-designed agreements not only protect individuals' privacy but also provide clarity and reassurance to all participating institutions.

3.9 Conclusion

Arranging partnerships and data access is the foundation of any MPD initiative. It requires a strong understanding of private-sector motives and concerns, rigor in legal and governance design, and strategic clarity about institutional roles and public value. By systematically assessing maturity, articulating incentives, defining boundaries, and formalising responsibilities, practitioners can create partnerships that are not only viable but sustainable over the long term.

In practice, every initiative will differ. The concepts outlined in this chapter provide an indication of how to navigate the complexity of establishing this data access in order to turn MPD into a reliable resource for policy and public good.

4 Chapter 4: Data Processing and Data Pipelines for MPD Initiatives

MPD initiatives rely on complex data processing and pipeline architectures to transform raw network events into robust, policy-relevant statistics. This chapter provides practitioners and decision-makers with a coherent, end-to-end description of different components of MPD pipelines; from generation and collection to processing, aggregation, and dissemination while at the same time focusing on the need to maintain data quality, privacy, and statistical validity.

4.1 Overview of the MPD Pipeline

An MPD pipeline can be understood as a sequence of interdependent stages that progressively transform raw network data into usable statistical outputs ([Ricciato et al. 2020](#); [Salgado et al. 2021](#)). While implementations differ across countries and institutions, most pipelines share five core stages: data collection, extract–transform–load (ETL), data cleaning and processing, aggregation and scaling, and dissemination. Each stage introduces specific technical, methodological, and governance considerations, and weaknesses at any point can compromise the integrity of the final results.

The pipeline begins with data generation and collection at the MNO, where individual phone activities create network events. These events are then prepared for analytical use through transformation and pseudonymisation. Subsequent processing stages add analytical value by correcting errors, inferring behaviour, and constructing meaningful indicators. Finally, aggregation and scaling convert processed data into population-level statistics suitable for publication and policy use ([Ricciato et al. 2020](#); [Salgado et al. 2021](#)).

4.2 Data Collection and Generation

MPD originates from routine interactions between subscribers and the mobile network. Each time a user makes a call, sends a message, or uses mobile data, the activity is routed through a specific network antenna or cell ([United Nations Statistics Division 2019](#); [Blondel et al. 2015](#)). The network operator records this interaction as an event, typically capturing three fundamental attributes: a subscriber identifier, a timestamp, and a location reference linked to the serving antenna.

Although this structure is simple, the scale is immense. Large operators generate billions of such records over relatively short periods. This volume underscores the importance of designing efficient downstream processes and minimising unnecessary data handling. Before data can leave the operational systems where it is generated, several preparatory steps are required. These include agreeing on a standardised data structure, defining formats for timestamps and identifiers, and ensuring that all parties share a common understanding of variable definitions. Without such alignment, downstream processing becomes error-prone and costly.

A critical element of this stage is pseudonymisation. Direct identifiers, such as phone numbers, are replaced with non-reversible codes so that individuals cannot be directly identified in subsequent processing stages. While pseudonymisation reduces risk, it does not eliminate sensitivity; MPD remains highly granular and requires strong safeguards throughout the pipeline.

4.3 Extract, Transform, and Load (ETL)

The ETL stage bridges raw data collection and analytical processing. During extraction, the agreed subset of network data is selected from operator systems. Transformation then reshapes this data into standardised formats, applies pseudonymisation, and derives additional attributes where necessary. For example, mobile country codes may be extracted to distinguish domestic subscribers from inbound roamers, which is essential for tourism analysis.

Data minimisation is a central principle at this stage. Given the sheer volume of records, unnecessary attributes should be removed, and efficient storage formats should be used ([Montjoye et al. 2018](#)). Data minimisation reduces transmission costs, lowers processing overhead, and limits exposure of sensitive information.

Secure transmission is another key concern. Data is typically transferred from the network operator to a processing environment, which may be internal to the operator or managed by a government body or service provider. Encryption, strong authentication, and integrity checks are essential to ensure that data is neither intercepted nor altered during transfer. In practice, large datasets are often partitioned into batches, by time period, geography, or subscriber group, to improve reliability and performance.

4.4 Data Cleaning: Ensuring Analytical Validity

Raw MPD contains numerous artefacts that can distort analysis if left unaddressed ([Blondel et al. 2015](#); [Salgado et al. 2021](#)). Data cleaning is therefore a foundational step in the pipeline, motivated by the principle that flawed inputs inevitably produce flawed outputs.

One common source of error is non-human activity. Internet-of-Things devices, such as vehicle trackers or security systems, generate network events but do not represent human behaviour. These records must be identified and removed when the objective is to measure population mobility. Other errors arise from the physical characteristics of networks and user behaviour. Phones used on aircraft or ships may connect to terrestrial antennas, creating the illusion that people are present on land when they are not. Similarly, accidental roaming near borders can cause subscribers to appear in a neighbouring country without having crossed it. Network artefacts can even produce “ghost events,” where a subscriber appears to be in two places simultaneously. Each of these phenomena requires explicit filtering rules.

Cleaning also serves a practical purpose by reducing data volume. Removing irrelevant or erroneous records lowers computational costs and improves the efficiency of subsequent processing stages.

4.5 Core Data Processing and Methodological Models

Once cleaned, data enters the core processing stage, where analytical value is created. A key design choice at this point is whether to adopt a centralised core data model or a decentralised, domain-specific approach. A centralised core model transforms cleaned event data into a common intermediate representation that supports multiple analytical domains, such as tourism, migration, or transportation. Domain-specific algorithms are then applied on top of this shared foundation. This approach promotes consistency and comparability across outputs. In contrast, decentralised models develop separate pipelines for each domain, which may be easier to implement initially but risk producing incompatible results over time.

Within a core model, several methodological techniques are commonly applied. Continuity models address the sparse nature of event data by inferring presence between observed activities ([Ricciato et al. 2020](#); [Salgado et al. 2021](#)). Rather than assuming that a person is present only at the moment of a recorded event, continuity models extend presence over plausible time intervals. This correction is essential for accurate population counts and exposure estimates.

Continuity models also enable the identification of “stays” and “moves.” Stays represent periods when an individual remains in a location, while moves capture transitions between locations. This distinction underpins analyses of commuting, travel behaviour, and transport mode inference. Another critical processing step is the detection of meaningful locations, such as home and work. By analysing spatial-temporal patterns such as where a subscriber spends most nights, analysts can infer habitual locations and define a person’s usual environment. Departures from this environment form the basis for identifying tourism trips and other forms of temporary mobility ([Isaacman et al. 2011](#)).

Accuracy can be further improved by integrating auxiliary data sources. Building footprints, land-use data, and road networks help constrain probabilistic location assignments, ensuring that inferred positions are plausible. These enhancements improve both spatial and temporal precision without increasing intrusiveness.

4.6 Aggregation and Indicator Construction

Processed data must be aggregated before it can be disseminated. Aggregation typically occurs along spatial and temporal dimensions, but the choice of units has profound implications for interpretation and privacy.

Spatial aggregation may align with administrative boundaries, such as municipalities or regions, or use regular or adaptive grids. Each approach involves trade-offs between interpretability, accuracy, and disclosure risk. Temporal aggregation similarly requires careful planning, as indicators may be produced at hourly, daily, monthly, or annual resolutions.

Beyond space and time, aggregation also reflects analytical intent. In tourism statistics, for example, analysts must decide whether to count visits, visitor-days, or nights spent, and whether to disaggregate by country of residence. These decisions should be made early, as they influence upstream processing requirements.

4.7 Scaling, Bias Adjustment, and Quality Assurance

MPD does not directly represent the total population. Some individuals carry multiple devices, while others may not use mobile phones at all. As a result, raw counts must be scaled and adjusted to correct for over-coverage and under-coverage. ([Flowminder Foundation 2023a](#); [Wesolowski et al. 2013](#); [Ricciato et al. 2020](#))

Scaling models often rely on operator market shares or external benchmarks to align mobile phone-based indicators with known population totals ([Wesolowski et al. 2013](#); [Cabrera and Rowe 2025](#); [Ricciato et al. 2020](#)). This step is methodologically challenging due to limited ground truth data, but it is essential for producing credible official statistics. Quality assurance accompanies scaling. Analysts must assess completeness, consistency, and plausibility, and apply statistical disclosure control techniques to prevent re-identification, particularly when publishing highly granular data.

4.8 Privacy by Design in the Pipeline

Privacy by design is operationalised by distinguishing between three tiers of data sensitivity. Tier 1 data consists of raw, identifiable records and remains under the strict control of the MNO. Tier 2 data is pseudonymised and used for processing under controlled conditions. Tier 3 data is fully aggregated and suitable for dissemination ([Montjoye et al. 2018](#)).

Understanding these tiers helps organisations design appropriate technical and organisational safeguards at each stage of the pipeline. Access controls, encryption, auditing, and strict role-based permissions are essential for Tier 1 and Tier 2 data, where risks of re-identification or commercial sensitivity are highest. Even at Tier 3, where data are aggregated and prepared for release, disclosure control remains necessary to ensure that small cell sizes, rare combinations of attributes, or extreme values do not inadvertently reveal information about individuals or commercially sensitive patterns. Privacy by design therefore operates as a continuous principle across the entire pipeline, rather than a single compliance step ([Montjoye et al. 2018](#)).

4.9 Pipeline Deployment Models

Finally, organisations must decide where processing occurs. In some models, MNOs perform all processing and provide only aggregated outputs to government users. In others, governments or trusted service providers process pseudonymised data within secure environments, sometimes hosted by the operator. Each model involves trade-offs between control, cost, capacity, and risk.

Regardless of the chosen arrangement, clarity of roles, responsibilities, and safeguards is critical. Successful MPD initiatives are characterised not only by technical sophistication, but by careful institutional design and sustained collaboration between data providers and users.

4.10 Conclusion

Effective planning for MPD initiatives requires a holistic understanding of data processing and pipelines. By viewing the pipeline as an integrated system, rather than a series of isolated steps, practitioners can make informed decisions that balance analytical ambition with feasibility, privacy, and quality. The principles outlined in this manual provide a foundation for designing robust, transparent, and policy-relevant MPD pipelines.

5 Chapter 5: Data Quality and Characteristics

This chapter revisits what CDRs are, what information they contain, why its characteristics matter for statistical quality, how common biases arise and can be adjusted, and how to operationalise quality assurance so that CDR-derived statistics are reliable and trusted for policy use ([Blondel et al. 2015](#); [Wesolowski et al. 2013](#); [Ricciato et al. 2020](#)).

5.1 How CDRs are generated and key data fields

As described in Chapter 1, CDR records are created when a subscriber makes or receives a call, sends or receives an SMS, and/or uses mobile data. Each record is associated with the network infrastructure that serviced the event; most importantly, the cell tower (cell ID) the device connected to at the moment of the event. A key implication is that the “location” in CDR is not the device’s precise coordinates; it is the tower location (or, more precisely, the tower identifier, which can be linked to coordinates via a separate reference table).

At a minimum, to use CDR data for producing useful outputs three fields will be needed:

1. **A Pseudonymised subscriber identifier** (the ‘hashed’ ID that removes personally identifiable information¹)
2. **A cell ID** (tower identifier),
3. **A timestamp** of the network event.

Many implementations also include additional metadata such as the receiving party identifier (also Pseudonymised) and the type of event (voice/SMS/data). Critically, CDR does **not** include communication content (no call audio, no SMS text); it contains metadata about the event.

¹Substituting the identifier through pseudonymisation protects identity by stripping out easily identifiable data such as a telephone number. Maintaining a stable pseudonymised identifier can still enable longitudinal analysis where this is lawful, necessary, and appropriately governed.

5.1.1 Pseudonymisation and identifiers: why they matter for quality

Pseudonymisation is described as replacing personally identifying values with random strings, with the important property that the **same original value maps consistently to the same Pseudonymised value**. This allows analysts to follow a device/subscriber across many records without exposing direct identifiers. Pseudonymised data should still be treated as sensitive, because repeated location patterns may enable re-identification when combined with other information.

CDR ecosystems also include multiple identifiers, and understanding them is useful because they can affect your unit of analysis (person vs SIM vs handset) and therefore your bias risks:

- **MSISDN** (the phone number) is tied to the subscriber and is used for routing and billing.
- **IMSI** is tied to the **SIM** and identifies the subscriber within the network.
- **IMEI** is tied to the **handset/device** (the physical phone).

CDR data should not be assumed to provide a direct record of individual persons. Records are typically associated with subscriber-, SIM-, device-, or subscription-level identifiers, which may not map one-to-one to people. From a quality standpoint, these distinctions become practical when dealing with realities like multiple SIM ownership, SIM swapping, or shared devices. Each of these can distort interpretations if we assume “one identifier equals one person” without testing that assumption.

5.2 Strengths and limitations of CDR data

5.2.1 Strengths: coverage, timeliness, and passive generation

CDR’s major advantages for policy analytics are that it can include **all subscribers** (regardless of device type or operating system) because it is collected primarily for billing and network management ([United Nations Statistics Division 2019](#); [Blondel et al. 2015](#)). It is generated passively and routinely collected as part of ordinary network operation, without the need to design a new data-collection instrument. It can also be near real-time and granular in the sense of high record volume.

5.2.2 Limitations: spatial precision, temporal intermittency, and representativeness

CDR quality constraints arise directly from how it is created:

- **Spatial precision is tower-dependent.** Location is approximated at cell-tower level, not precise coordinates. Tower density varies, so spatial precision is typically much higher in urban areas and lower in rural areas.

- **Temporal resolution is behaviour-dependent.** CDR observations occur when network events occur. A user who rarely calls/texts/uses data will have sparse observations, producing intermittent trajectories (unlike many GPS streams that are logged at regular intervals).
- **Representativeness is not guaranteed.** Not everyone owns a phone, not everyone is on the same operator, and not everyone uses their phone in the same way. Demographics and socioeconomic factors influence who appears in the data and how. Note also that one subscription does not always equal one individual (shared phones, multiple SIMs per person).

A useful way to internalise these limitations is to treat CDR as a high-volume observational dataset whose measurement properties vary across geography, network configuration, and subpopulations, rather than as a direct census of “people” or “movements” ([Blondel et al. 2015](#); [Wesolowski et al. 2013](#); [Ricciato et al. 2020](#)).

5.3 Interpreting location in CDR: from records to meaningful places

5.3.1 Mobility as “movement between towers,” not continuous location traces

CDR-based mobility is observed as changes in connected tower IDs recorded in network events over time. For example, a day’s records might show a device associated with tower A in the early morning, tower B during commuting and working hours, tower C during mid-day, and tower A again at night, suggesting a simple sequence of tower-to-tower transitions. This is conceptually different from GPS-derived data, which may provide more frequent and detailed location traces, depending on how the data are collected.

5.3.2 Using timestamps to infer home and work

CDR does not label locations as “home” or “work.” These locations are usually inferred from repeated time-of-day patterns, based on the assumption that people are typically at home at night, away during working hours, and return in the evening. Therefore, towers observed late at night may be used as proxies for home, while towers observed consistently during working hours are candidates for workplace location. This inference is not perfect, but it is a pragmatic, explainable method grounded in time-of-day behaviour ([Isaacman et al. 2011](#)). To apply this responsibly, it is generally necessary to:

- Define time windows carefully (e.g., “night” vs “work hours”) and test sensitivity.
- Require sufficient observation volume (sparse users may not support reliable inference).
- Validate against reference data where available (surveys, travel studies, or known anchor points).

5.4 Spatial coverage and the Voronoi approximation: what it enables and what it hides

5.4.1 Voronoi cells as a workable model of tower coverage

The **Voronoi diagram** is a useful tool for approximating spatial coverage, partitioning space into regions where each point is closest to a given tower (Ricciato et al. 2020). Each polygon (“Voronoi cell”) can serve as a practical unit for associating CDR events to geographic areas. Note that Voronois are only a useful simplification of reality. The true network coverage will depend on factors such as terrain, capacity, and network configuration. But Voronoi cells provide a useful baseline for analysis in the absence of more detailed or nuanced network coverage data.

5.4.2 Why CDR trajectories can deviate from “actual mobility”

Note that flows among towers can deviate from real-world movement paths. Two reasons for this are:

1. **Variable cell size:** urban areas have many towers, smaller cells, and therefore finer spatial precision; rural areas can have large cells where within-cell movement is invisible.
2. **Sparse observation:** if a device generates few events, you will only observe fragments of movement, and tower-to-tower transitions will appear as “straight jumps” that omit intermediate routes.

Therefore, CDR data is often excellent for measuring *aggregate flows and population dynamics* at suitable spatial-temporal scales, while reconstructing precise routes or micro-mobility in low-density tower environments usually requires supplementary data, stronger assumptions, or additional modelling (Calabrese et al. 2011; Toole et al. 2015; Deville et al. 2014).

5.5 Key data-quality considerations driven by network behaviour

5.5.1 Tower density and the visibility of short trips

Tower density is closely related to population density because operators deploy more infrastructure where there are more users and higher demand (Blondel et al. 2015; Ricciato et al. 2020). In urban areas, the high tower density makes it more likely that short trips (e.g., within a neighbourhood) will register as changes in connected towers. In rural areas, short trips may remain entirely within one tower’s coverage and therefore appear as “no movement.”

This matters because many downstream indicators (commuting intensity, localised displacement, neighbourhood-level service catchments) are highly sensitive to the minimum observable movement unit. If tower-density effects are not explicitly accounted for, the analysis may inadvertently under-detect mobility in rural or low-coverage regions.

5.5.2 Handover and location noise: false mobility signals

One common source of location noise is frequent tower switching (handover), which can occur due to fluctuating signal strength, load balancing, or switching among technologies (Blondel et al. 2015; Salgado et al. 2021). The key risk is that CDR may show multiple tower changes even when a person is essentially stationary, creating false mobility signals, especially where tower coverage overlaps.

One possible mitigation approach is to **cluster nearby towers**: when towers are very close and frequent switching occurs among them, treating them as a single “location cluster” can reduce false movement and produce more stable mobility signals.

In practice, clustering is not just a technical adjustment; it is part of making the measurement model explicit. This means that, at the spatial scale of interest, switching among certain towers should not be interpreted as meaningful movement.

5.6 Representativeness and bias: why “who is in the data” is central to quality

CDR-based analysis observes only a subset of the population:

- People with **no subscription** are absent.
- People subscribed to **other operators** are absent unless you have multi-operator coverage.
- Even within a given operator, **inactive subscribers** contribute little or no usable data because they rarely generate events.

Survey data from Ghana has shown the differential phone use across groups, for example, differences by gender, age, and urban/rural locality, illustrating how underrepresentation can arise even when overall penetration is high. From a statistical-quality perspective, this creates a classic problem: the dataset is not a random sample of the target population. Instead, it is shaped by access (ownership), operator market share, and usage intensity, all of which correlate with demographic and socioeconomic factors (Wesolowski et al. 2013; Cabrera and Rowe 2025). The result is **systematic bias** unless addressed explicitly.

For practical quality assessment, representativeness bias can be organised into several layers, each reflecting how the observed CDR population may differ from the target population:

- Population coverage bias: people without access to mobile phones or subscriptions are not represented in the data.

- Operator coverage bias: subscribers of non-participating operators are not observed, and market share may vary by region or socioeconomic group.
- Activity bias: even among subscribers of participating operators, people who do not use their phones frequently generate fewer records and are less visible in the data.
- Identifier bias: one subscription, SIM, or device does not necessarily correspond to one person, due to multiple SIM ownership, shared phones, or device changes.

5.7 Adjusting for bias: population-weighted adjustment and triangulation with other sources

Population-weighted adjustment is one approach to improving representativeness by combining CDR with other population data such as censuses, gridded population estimates, or surveys ([Wesolowski et al. 2013](#); [Cabrera and Rowe 2025](#); [Ricciato et al. 2020](#)). The basic workflow is:

- Estimate residential population associated with towers using CDR and aggregate to an administrative unit (e.g., district).
- Compute a scaling factor per unit as (reference population) / (CDR-estimated population).
- Multiply CDR-derived estimates by these scaling factors to align them with population benchmarks.

Adjustments can be refined depending on what auxiliary data exists. For example, information about known user/non-user patterns, multiple-SIM ownership, or differences across operators. This framing is practical: bias correction is not a single formula, but a set of strategies that should be tailored to the country context and the specific indicator.

Multiple sources can be combined in real projects (e.g. baseline gridded estimates such as WorldPop produces can be used in some contexts, national census in others, together with secondary and primary surveys). The point here is that **CDRs alone are not enough**; quality improves when CDR data gets triangulated with established statistical sources.

Bias adjustment improves representativeness but does not remove all uncertainty. Because scaling factors depend on reference data quality and modelling assumptions, key assumptions should be documented, sensitivity checks conducted, and adjusted outputs validated against independent data sources where feasible. ([Flowminder Foundation 2023b](#))

5.8 Quality assurance and trust: moving from analysis to statistics suitable for policy use

5.8.1 Why quality assurance is not optional

When CDR is used for policy, maintaining **public trust** is essential, and that trust is not only about privacy (see also Chapter 6). If outputs are not reliable and well-documented, public trust and user confidence erodes even if security is strong.

5.8.2 A structured quality assurance approach: ESS QAF ([European Statistical System 2019](#)) and “quality gates”

A quality assurance framework has been developed by the European Statistical System. This is a collection of good practices and tools applicable at institutional, process, and output levels. For big-data projects, quality assurance can be organised around stages that align with the concept of the data pipeline: **input**, **throughput (processing)**, and **output**. ([Ascari et al. 2024](#)) In the context of CDR, input quality concerns the completeness, consistency, and statistical usability of the original event records and network reference data. Throughput quality covers the processing steps that transform these data into indicators, including cleaning, validation, integration, scaling, and aggregation. Output quality focuses on whether the resulting statistics are relevant, coherent, comparable, and accompanied by clear documentation of assumptions and limitations. A particularly useful concept here is the use of **three quality gates**, each acting as a checkpoint before proceeding.

Quality Gate 1: Input data (raw acquisition and validation). At the start of the pipeline, review raw data as received from operators, check consistency, perform logical checks, and confirm validity for statistical use. If faults are detected, do not proceed. Request correction from the operator, since many issues (format changes, missing fields, inconsistent logs) are best resolved at source.

Quality Gate 2: Throughput (processing design, testing, and stabilisation). During system design, multiple parameters and settings must be tuned. The training describes iterative “test loops” where alternative settings are run, processing logs are reviewed, and methods/algorithms are refined until the system produces stable, expected outputs. This gate is less about a single pass/fail rule and more about disciplined iteration with documented decisions.

Quality Gate 3: Output (statistical validation and coherence). At the output stage, review results for accuracy, consistency, comparability, and coherence. Check that output tables align logically, and validate against reference data where feasible. One real challenge for big data is that some phenomena captured well by CDR (e.g., short-term mobility, daytime population) may lack direct reference datasets for validation, so validation may need to focus on specific geographies, periods, or subsets where comparisons are possible.

5.9 Conclusion

Below summarizes key considerations for maintaining data quality in MPD initiatives:

- **Treat CDR location as a measurement model, not a ground truth.** Because CDR records the tower connection (and not device coordinates), “location” is a proxy with geography-dependent error. The analytical decisions such as tower-to-area mapping often using Voronoi cells, clustering, spatial aggregation, are part of a measurement model and should be documented and justified.
- **Make temporal assumptions explicit and test them.** Many inferences (home/work, commuting, exposure) depend on time windows and routines. These can work well, but only when backed by sufficient observation density and sensitivity analysis. Sparse users and irregular schedules can otherwise introduce systematic misclassification.
- **Control for network artifacts (handover) before interpreting mobility.** If you do not address handover noise, you will systematically overestimate movement in overlapping-coverage regions. Clustering proximate towers is a practical mitigation, but it must be calibrated to the required spatial resolution and validated for stability.
- **Interrogate representativeness at multiple layers: population, operator, and activity.** MPD typically covers only (a) people with phones, (b) on the participating operator(s) network, (c) who are active users of their phones. Each filter can introduce demographic and geographic bias. This is not a minor caveat; it is a core quality dimension for policy statistics.
- **Plan for bias adjustment and validation from the start.** Population-weighted adjustment and triangulation with census/survey/gridded sources should not be an afterthought. They are central to making outputs interpretable and defensible. Projects should be designed in such a way that reference data can be integrated early. The assumptions and scaling factors that underpin analysis should also be made clear and shared transparently.
- **Operationalise quality with gates, logs, and reproducible processes.** The three quality gates provide a workable governance structure: validate inputs, stabilise processing, and verify outputs. This structure supports not only technical correctness but also auditability, an essential property when results are used for decisions and public communication.

6 Chapter 6: Data Governance and Safeguards in MPD Initiatives

The processing of data within an MPD initiative must be conducted in a legal, ethical, and equitable manner, with due regard for the interests of all affected stakeholders. Because MPD initiatives rely on data generated through the everyday use of mobile networks, they raise distinct governance challenges related to privacy, security, commercial sensitivity, and public trust ([Jansen et al. 2021](#); [Montjoye et al. 2018](#)). These challenges are not peripheral to technical implementation; they are foundational to whether an initiative is lawful, credible, and sustainable.

This chapter provides a structured overview of data governance as it applies to MPD initiatives. It explains what it means to process MPD, identifies the relevant stakeholders and their interests, clarifies the distinction between personal and non-personal data, and outlines the principal legal, ethical, and security risks. It then introduces practical safeguards and principles that support the responsible and sustainable use of MPD for public policy and statistical purposes. Introductory courses to data governance and online resources are signposted in Appendix 1. ([International Telecommunication Union n.d.b](#); [FlowGeek n.d.](#))

6.1 What Does It Mean to Process Data?

Data processing encompasses a wide range of activities, including the collection, storage, transformation, analysis, sharing, and dissemination of data. In the context of MPD initiatives, processing most often involves handling data originally collected by MNOs for operational and billing purposes, such as CDRs. Although these data are not collected for analytical or policy use, any subsequent use constitutes processing and is therefore subject to legal and ethical constraints.

Processing becomes particularly sensitive when it involves personal data or data derived from personal data ([Montjoye et al. 2013, 2018](#)). In addition to privacy considerations, MPD initiatives may also involve commercially sensitive information, national infrastructure data, or information relating to vulnerable populations. Effective data governance must therefore address a broad range of sensitivities and risks, not only those associated with individual privacy.

6.2 Data Governance and the Stakeholder Ecosystem

Data governance refers to the policies, processes, and controls that ensure data are used responsibly, securely, and in compliance with applicable laws and ethical standards. In MPD initiatives, governance is inherently multi-stakeholder, reflecting the fact that data originate in the private sector, are often processed for public purposes, and affect individuals and society at large.

Key stakeholders typically include mobile phone subscribers, MNOs, regulators, data users such as NSOs or government ministries, and civil society. These stakeholders have different, and sometimes competing, priorities. Subscribers are primarily concerned with privacy and protection from misuse. Operators must safeguard both subscriber trust and commercially sensitive information. Regulators are responsible for enforcing legal compliance and protecting public interests. Data users require data that are accurate, timely, and fit for purpose. Civil society has a broader interest in ensuring that MPD is not misused in ways that undermine rights, equity, or public trust.

Effective governance requires acknowledging these differing interests and establishing mechanisms to balance them. This includes clearly defining roles and responsibilities, ensuring transparency in how data are used, and implementing safeguards that address both privacy and commercial concerns.

6.3 Personal and Non-Personal Data in MPD Initiatives

A central task in data governance is determining whether data qualify as personal data. While definitions vary across jurisdictions, personal data are generally understood to be any information relating to an identified or identifiable natural person. Identification may be direct, such as through a name or phone number, or indirect, where identification becomes possible through combination with other data.

In MPD initiatives, this distinction is particularly important because mobility data are inherently identifying. Even when explicit identifiers such as phone numbers or subscriber IDs are removed or replaced, individual movement patterns are often unique and highly regular. As a result, individual-level mobility data remain personal data, regardless of whether direct identifiers are present. Removing names or numbers alone does not anonymise such data ([Gonzalez et al. 2008](#); [Song et al. 2010](#); [Montjoye et al. 2013](#)).

Non-personal data, by contrast, do not relate to any identifiable individual. In practice, most MPD initiatives rely on aggregated data products that summarise patterns across large groups of subscribers, rather than individual trajectories. However, whether data are genuinely non-personal depends on the level of aggregation, the availability of auxiliary information, and the evolving state of reidentification techniques. Governance frameworks must therefore adopt a cautious and context-aware approach to classification, recognising that what is considered anonymised today may not remain so in the future ([Montjoye et al. 2013, 2018](#)).

6.4 Legal and Regulatory Context

MPD initiatives operate within complex legal and regulatory environments that vary across jurisdictions. Most countries now have some form of data protection or privacy legislation, and many have sector-specific regulations governing telecommunications data, cybersecurity, or national security. These frameworks determine who may process personal data, for what purposes, under what conditions, and with what safeguards.

For training purposes, the General Data Protection Regulation (GDPR) ([GDPR.eu n.d.](https://gdpr.eu)) provides a useful reference point, as many jurisdictions have adopted similar principles. Core concepts include lawfulness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality, and accountability. While the specific legal obligations differ by context, these principles offer a broadly applicable framework for thinking about responsible data use.

Importantly, there is no single governance model that applies universally. Legal obligations depend not only on the country where data originate, but also on where data are processed and who is involved. MPD initiatives must therefore be grounded in jurisdiction-specific legal analysis, supported by engagement with relevant regulators and legal experts.

6.5 Risks Associated with MPD Initiatives

MPD initiatives entail a range of interrelated risks that must be proactively identified and mitigated. Privacy risks include unauthorised access to sensitive data, reidentification of individuals, and the use of data for surveillance or profiling. Security risks encompass data breaches, whether through malicious attacks, inadequate access controls, or accidental disclosure. Ethical risks arise when data are misused, misinterpreted, or applied in ways that exacerbate bias, exclusion, or harm to vulnerable populations ([Montjoye et al. 2013, 2018](#)).

These risks are not hypothetical. Experience shows that even well-intentioned initiatives can undermine public trust if governance arrangements are weak or poorly communicated. Risk management must therefore be treated as an ongoing process rather than a one-time compliance exercise.

6.6 Safeguards and Mitigation Measures

Mitigating governance risks requires a combination of technical, organisational, and procedural safeguards. From a privacy perspective, common measures include Pseudonymisation, aggregation, and redaction. Pseudonymisation replaces direct identifiers with random values, allowing records to be linked without revealing identities. Aggregation summarises data across space and time, reducing the visibility of individual behaviour. Redaction techniques, such as enforcing minimum group sizes, help prevent disclosure in sparsely populated areas or time periods.

These measures must be applied thoughtfully, balancing privacy protection against data utility ([Montjoye et al. 2018](#)). Stronger safeguards generally reduce analytical precision, making it essential to align protection levels with clearly defined purposes.

Security measures are equally critical. Sensitive data should be stored behind secure firewalls, encrypted at rest and in transit, and accessed only by authorised personnel under strict access controls. Wherever possible, anonymisation and aggregation should occur within the secure environments of operators or regulators, minimising data movement and exposure. Logging, auditing, and regular security reviews further strengthen accountability.

6.7 Best Practices for Protecting Privacy in CDR analysis

6.7.1 Current, standard methods for preserving individual privacy

Personal data must have strong protections to preserve the privacy of individuals in the dataset. Current practice usually has three types of privacy-preserving methods applied at different stages of the data pipeline:

- Pseudonymisation
- Aggregation
- Redaction

6.7.1.1 Pseudonymisation

Pseudonymising data involves replacing directly identifying information with randomly-generated values, such that the linkage between records is preserved (i.e. identical values remain identical) ([Montjoye et al. 2018](#)). For MPD, this might involve replacing subscribers identifiers (e.g. phone number, IMSI, MSISDN) with a random string.

Known as ‘hashing’ this process removes the personal information which allows the direct identification of subscribers, whilst still enabling the records from an individual subscriber to be linked together in order to analyse mobility.

As a result, pseudonymisation obscures subscribers’ identities but may not anonymise the data as individuals may still be identified from their mobility patterns which are maintained in this process ([Montjoye et al. 2013](#)).

6.7.1.2 Aggregation

By combining or aggregating MPD data for very large numbers of subscribers we can help preserve their privacy by making the movements of any single subscriber difficult, if not impossible, to discern. However, we still can make inferences about the distribution and mobility of the population as a whole from aggregated data.

Data is aggregated spatially (e.g. by district or region) and temporally (e.g. by day, month) at a resolution informed by the type of indicator being produced and the requirements of the data user. For example, MPD for population estimates may be spatially aggregated at a district-level resolution and temporally at a monthly resolution, while transport applications would require higher spatial and temporal resolution aggregates.

In addition to helping protect the individual privacy of the subscribers, the spatial aggregation of CDR data may remove other sensitive information about the number and locations of cell towers in each area which may be a concern of other stakeholders such as the operator.

However, aggregation may not be sufficient to protect the individual privacy of all subscribers. Aggregation relies on there being a sufficient number of subscribers in each area in each time frame to prevent any individual being reidentified. Without any further checks, only aggregating CDR data risks producing outputs in which there is only a single or very few subscribers in a given location at a given time which may risk their reidentification. This is more likely to occur at high spatial and temporal resolution ([Montjoye et al. 2013](#)).

6.7.1.3 Redaction

To better ensure that the individual privacy of subscribers is preserved, we can use additional anonymisation frameworks such as k-anonymity ([Montjoye et al. 2013, 2018](#)). A dataset can be described as k-anonymised if each subset of data points for a given individual (i.e. their location at each time point) is shared by at least k-1 other subscribers. For example, if we set k to 15 this means that the aggregated CDR must have at least 16 subscribers associated with (present or resident in) each location at each time point. Any combinations of location and time associated with fewer than 15 subscribers are redacted from the data set.

While ensuring k-anonymity with a suitable threshold is currently sufficient to preserve the individual privacy of subscribers in a CDR dataset, anonymisation is a moving target as new methods for reidentification of subscribers and for data protection continue to be developed ([Montjoye et al. 2018](#)).

6.7.2 Emerging and novel Privacy-Enhancing Technologies (PETs)

Researchers continue to develop new tools and techniques to preserve individual privacy. Extensions of k-anonymity such as [historical k-anonymity](#) and [L-diversity](#) have been proposed to provide further protection, particularly for GPS data which is more vulnerable than CDR data to reidentification attacks due to the greater spatial and temporal resolution. [Privacy researchers](#) have also proposed using neural networks to generate synthetic mobility datasets. These are based on real datasets and maintain the same aggregated statistical properties and patterns, but are not generated directly from the mobility data of real subscribers, meaning no subscribers can be reidentified. However, these techniques are still being developed and are not currently necessary for the anonymisation of MPD.

When considering the appropriate implementation of privacy-preserving methods, and especially the resolution that MPD is aggregated at, it is important to recognise the trade-off with data quality.

6.8 International frameworks and industry standards

Existing international frameworks and industry standards can be a useful way of informing the sustainable and ethical use of MPD. Examples of these include the guiding principles developed by the UN-CEBD MPD task team for maintaining public trust when using MPD; the Locus Charter which has identified a set of ethical guidance principles for any type of location data and the GSMA's Mobile Privacy Principles.

6.8.1 UN Guiding Principles for maintaining public trust when using MPD

Whilst MPD can be useful for multiple applications (see Chapter 2), public trust cannot be assumed; it must be actively maintained through clear standards, transparent practices, and professional accountability. To support this, the UN-CEBD MPD task team developed a set of five guiding principles to frame the responsible use of mobile operator data in policy contexts, particularly by public institutions and national statistical systems. ([Jansen et al. 2021](#); [United Nations 2014](#)) The five key principles for maintaining public trust when using MPD are:

1. **Necessity and Proportionality:** Mobile operator data should only be used where there is a clearly defined public policy need and where existing data sources are insufficient. The scope, granularity, and frequency of data use should be limited to what is strictly necessary to achieve the stated policy objective.
2. **Professional Independence:** The production and interpretation of indicators derived from mobile operator data should be carried out independently of political or commercial influence. Transparent methods and clear documentation are essential to ensure credibility and accountability, particularly when outputs inform public decision-making.
3. **Privacy Protection:** Strong safeguards must be applied to prevent identification of individuals, including aggregation, anonymisation, and the use of minimum thresholds. Compliance with applicable data protection laws and clear communication about privacy measures are critical to sustaining public trust.
4. **Commitment to Quality:** Outputs based on mobile operator data should meet standards comparable to official statistics, including accuracy, consistency, timeliness, and transparency about limitations or biases. Quality assurance processes should be embedded throughout data processing and analysis.
5. **International Comparability:** Where feasible, methods and indicators should be harmonised across countries to enable meaningful comparison and shared learning. Alignment with international statistical standards enhances the broader policy value and legitimacy of these data products.

6.8.2 The Locus Charter

The Locus Charter, (Benchmark Initiative 2021) launched in 2021, is a set of proposed common international ethical principles to help users of location data, including MPD, to make informed and responsible decisions. The Locus Charter proposes that “wider, shared understanding of risks and solutions relating to uses of location data can improve standards of practice, and help protect individuals and the public interest”. It has ten principles.

i Box 9: The 10 principles of the Locus Charter

Realise opportunities – Location data offers many social and economic benefits, and these opportunities should be realised responsibly. Understand impacts – Users of location data have a responsibility to understand the potential effects of their uses of data, including knowing who (individuals and groups) and what could be affected, and how. That understanding should be used to make informed and proportionate decisions, and to minimise negative impacts. Do no harm – Physical proximity amplifies the potential harms that can befall people, flora and fauna. Data users should ensure that the individual or collective location data pertaining to all species should not be used to discriminate, exploit or harm. Rights established in the physical world must be protected in digital contexts and interactions. Protect the vulnerable – Vulnerable people and places can be disproportionately harmed by the misuses of location data, and may lack the capacity to protect themselves. In these contexts, data users should take additional care, act proportionately, and positively avoid causing harm. Address bias – Bias in the collection, use, and combination of location datasets can either remove affected groups from mapping that conveys rights or services, or amplify negative impacts of inclusion in a dataset. Therefore care should be taken to understand bias in the datasets and avoid discriminatory outcomes. Minimise intrusion – Given the intimate and personal nature of location data, users should avoid unnecessary and intrusive examination of people’s lives and the places they live in, that would undermine human dignity. Minimise data – Most business and mission applications do not require the most invasive scale of location tracking available in order to provide the intended level of service. Users should comply with practices that adhere to the data minimisation principle of using only the necessary personal data that is adequate, relevant and limited to the objective, including abstracting location data to the least invasive scale feasible for the application. Protect privacy – Tracking the movement of individuals through space and time gives insights into the most intimate aspects of their lives. In the rare cases when aggregated and anonymised location data will not meet the specific business or mission need, location data that identifies individuals should be respected, protected, and used with informed consent where possible and proportionate. Prevent identification of individuals – As an individual’s mobile location data is situated within more and more geospatial context

data, its anonymity erodes, measures should be put in place to prevent subsequent use of the data resulting in identification of individuals or their location. Provide accountability – People who are represented in location data collected, combined, and used by organisations should be able to interrogate how it is collected and used in relation to them and their interests, and appeal those uses proportionate to levels of detail and potential for harms.

6.8.3 GSMA Mobile Privacy Principles

In 2011 the GSMA published Mobile Privacy Principles ([GSMA 2016](#)) that provide a widely recognised, globally applicable set of standards to guide how personal data is handled in the delivery of mobile services and applications. They noted that protecting user privacy has become essential to maintaining consumer trust and ensuring the long-term legitimacy of the mobile ecosystem. The principles established a common, user-centric framework intended to guide MNOs, application developers, and other ecosystem participants in the responsible collection, use, and sharing of personal data. The principles are designed to be applicable across jurisdictions, supporting consistent privacy practices while allowing flexibility to reflect local legal and regulatory requirements. The core principles are as follows:

1. **Transparency and Notice:** Organisations should provide clear, accessible information to users about what personal data is collected, how it is used, who it is shared with, and for what purposes. Transparency is essential to enabling informed user engagement and sustaining trust.
2. **User Choice and Control:** Users should be provided with meaningful choices regarding the collection and use of their personal data. Where appropriate, mechanisms should exist for users to grant, withhold, or withdraw consent and to manage their privacy preferences over time.
3. **Data Minimisation and Purpose Limitation:** The collection and retention of personal data should be limited to what is necessary, relevant, and proportionate to deliver the stated service or meet legitimate business or legal objectives. Data should not be used in ways that are incompatible with the original purpose without appropriate safeguards.
4. **Security Safeguards:** Appropriate technical and organisational measures should be implemented to protect personal data against unauthorised access, disclosure, alteration, or loss. The level of protection should reflect the sensitivity of the data and the risks associated with its use.
5. **Accountability and Governance:** Organisations should be accountable for their data practices and able to demonstrate compliance with applicable privacy principles and laws. This includes embedding privacy considerations into product design, internal policies, and operational processes.

These are supported by supplementary guidance, including privacy-by-design recommendations and accountability frameworks, intended to assist organisations in translating high-level principles into operational practices. Together, these resources promote consistent privacy standards, support regulatory compliance, and help ensure that innovation in mobile services does not undermine individual rights or public trust.

6.9 Conclusion

Data governance, including safeguarding and ethical use, is not an ancillary component of MPD initiatives; it is central to their legitimacy and effectiveness. By clearly defining purposes, understanding the legal and ethical context, engaging stakeholders, and implementing robust safeguards, organisations can responsibly harness MPD for public benefit. Governance frameworks must remain adaptive, reflecting changes in technology, regulation, and societal expectations. When treated as a core design consideration rather than a constraint, data governance enables MPD initiatives to deliver meaningful insights while protecting the rights and interests of all stakeholders.

7 Chapter 7: Managing the Communications Aspects of MPD Initiatives

Effective communication is a critical enabling factor for the success, legitimacy, and sustainability of MPD initiatives. Because these initiatives rely on sensitive data sources, involve multiple institutional actors, and often address issues of public interest, the way they are communicated can determine whether they are trusted, understood, and ultimately supported.

This chapter provides guidance on how to plan, manage, and deliver communications for MPD initiatives. It builds on general good practice in public-sector communication while addressing challenges that are specific to MPD, such as privacy concerns, technical complexity, and commercial sensitivity.

It is organised around three core dimensions: foundational communication principles, communication practices tailored to MPD initiatives, and operational considerations for managing communications within partnerships and public-facing contexts.

7.1 Foundational Communication Principles

Although MPD initiatives have unique characteristics, they are governed by a set of general communication principles that apply to most complex, multi-stakeholder projects. Applying these principles consistently helps ensure that communications are clear, credible, and fit for purpose.

7.1.1 Knowing and Segmenting the Audience

The first and most important principle is to understand who the communication is for. MPD initiatives typically engage a wide range of audiences, including technical experts, policy-makers, operators and regulators, civil society organisations, journalists, and the general public. Each of these audiences has different levels of technical knowledge, different interests, and different concerns.

Effective communication therefore requires deliberate audience segmentation and message tailoring. A technical briefing for telecom engineers may reasonably include references to network infrastructure, timestamps, and metadata, while a public-facing explanation of the same work should focus on high-level concepts and social value, using plain language and concrete examples. The underlying facts remain the same, but the framing, terminology, and level of detail must change.

Stakeholder mapping tools, such as influence–interest matrices, can help teams identify which stakeholders require close engagement, which need to be kept informed, and which may only require periodic monitoring. Using such frameworks supports strategic prioritisation of communication efforts and resources.

7.1.2 Avoiding Jargon and Explaining Acronyms

MPD initiatives sit at the intersection of telecommunications, statistics, and data science, fields that are rich in acronyms and specialised terminology. While such language is efficient among experts, it can quickly alienate or confuse non-specialist audiences.

As a general rule, acronyms should be spelled out on first use, and technical terms should be explained in clear, accessible language. Where possible, everyday alternatives should be preferred over specialised jargon. This does not mean sacrificing accuracy, but rather translating complexity into language that can be understood by the intended audience.

Clear communication begins with a simple but essential question: who is this message for, and what do they need to understand in order to engage meaningfully with the initiative?

7.1.3 Starting with the “Why” and the Public Value

Another core principle is to lead with purpose. Communications should start by explaining why the work is being undertaken, what problem it is intended to address, and why it matters.

Technical descriptions of data pipelines, models, or indicators should never be the entry point for non-technical audiences. Instead, communications should foreground the public value of the initiative, such as improving disaster response, supporting health planning, or strengthening transport systems and only introduce technical detail when it serves to deepen understanding.

When people understand the motivation and benefits of an initiative, they are more likely to trust it, support it, and engage constructively with its outputs.

7.1.4 Transparency and Honesty

Transparency is especially important in MPD initiatives because they involve data that people may perceive as personal or intrusive. Communications should be open and honest about who is involved, what data is being used, what safeguards are in place, and what the data can and cannot be used for.

Overly vague or defensive communication can create suspicion, while proactive and clear explanations help build confidence. Transparency does not require sharing sensitive details, but it does require explaining governance arrangements, legal bases, and accountability mechanisms in an accessible way.

7.1.5 Allocating Dedicated Communication Resources

Finally, effective communication requires deliberate investment. Communications should not be treated as an afterthought or an add-on to technical work. Successful initiatives typically assign clear responsibility for communications, allocate sufficient time and budget, and integrate communication planning into the overall project design, from the start.

A small but well-coordinated team, often combining project management, technical expertise, and communications skills, can ensure that outputs are accurate, coherent, and appropriate for their intended audiences.

7.2 Communication Principles Specific to MPD Initiatives

Building on general communication good practice, MPD initiatives require additional care because of the sensitivity of the data involved and the diversity of institutional interests.

7.2.1 Clarifying What Is Sensitive

Sensitivity in MPD work is context-dependent. What is considered sensitive information may vary across organisations, projects, countries, and regulatory environments. It is therefore essential to discuss and agree on sensitivities early in the initiative.

Two broad categories of sensitive information commonly arise. The first is personal data, such as individual-level CDR trajectories, which can reveal patterns of movement even when direct identifiers are removed. The second is commercially sensitive information, such as detailed network infrastructure data, which MNOs may need to protect for competitive or security reasons ([Montjoye et al. 2013, 2018](#)).

Clear internal agreements on what is sensitive, why it is sensitive, and how it will be handled are a prerequisite for responsible communication. Being transparent internally and, where appropriate, externally about these boundaries helps manage expectations and build trust among partners and stakeholders.

7.2.2 Explaining Data Sources and Safeguards Accessibly

Many audiences are unfamiliar with mobile operator data and may have misconceptions about what it contains. Communications should therefore include simple, accurate explanations of key data sources, such as CDRs, emphasising what they are and what they are not.

It is particularly important to explain data protection measures, such as pseudonymisation performed by operators and aggregation of results to population-level indicators. Explicitly stating that content of calls or messages is never accessed can help address common concerns and prevent misunderstanding ([Montjoye et al. 2018](#)).

The aim is to provide enough information to support informed engagement without overwhelming the audience with unnecessary technical detail.

7.2.3 Using Careful and Responsible Language

The language used to describe MPD initiatives shapes how they are perceived. Certain terms such as “tracking individuals” can trigger alarm or misrepresent the nature of the work. More accurate alternatives, such as “analysing population mobility patterns” or “producing aggregated insights,” better reflect both the methods and the intent of the analysis.

Careful language choice demonstrates respect for the audience and reduces the risk of misinterpretation by media or other third parties. Consistency in terminology across communications is also important, particularly in long-running initiatives.

7.2.4 Reiterating the Purpose and Benefits

Given the complexity of such initiatives, it is rarely sufficient to explain the purpose only once. Communications should repeatedly and consistently return to the “why,” reinforcing the public value and intended benefits of the work.

This repetition helps ensure that audiences retain the core message, even as technical details or specific results are introduced over time.

7.3 Transparency, Independence, and Public Engagement in Statistical Contexts

For MPD initiatives led by or involving NSOs, additional communication considerations apply.

Proactive communication with the public and key stakeholders at the outset of an initiative has been shown to foster more supportive reactions than reactive communication after concerns arise. Early engagement allows questions to be addressed before narratives of secrecy or misuse take hold.

Possible communication channels include public statements about statistical modernisation efforts, briefings to supervisory statistical councils, and staged dissemination of results as outputs mature. Decisions about timing and format should be informed by national context and stakeholder expectations.

At the same time, statistical agencies must preserve their professional independence. While methods and processes may be developed collaboratively with partners, the release and interpretation of official indicators should remain the responsibility of the statistical authority. Transparency about methods and governance arrangements supports this independence and strengthens credibility.

In addition, because the MPD analysis industry is still relatively new and new methods are continually being developed, the analytical methods used to produce useful information for policy makers and decision makers may not yet be adopted as standard methodologies to follow within the statistical community. That does not, however, necessarily preclude their use - it is just very important to communicate clearly both how the data was analysed and the extent to which the results may be deemed 'experimental'.

For this reason, some NSOs have chosen to publish 'experimental statistics' alongside their 'official statistics'. Doing so, creates the opportunity for decision-makers to use operational data that could be valuable to them in a setting such as responding to emergencies. Ensuring effective communication, therefore, about the nature of the data and the extent to which it is experimental is an important consideration.

7.4 Balancing Rigour, Accessibility, and Uncertainty

Communicating outputs derived from MPD requires balancing scientific rigour with accessibility. For non-specialist audiences, clarity and interpretability should take precedence over technical completeness, while still maintaining accuracy.

It is also essential to communicate uncertainty. Like all statistical outputs, indicators derived from MPD are estimates rather than exact counts. Explaining uncertainty, confidence intervals, and limitations in plain language helps prevent overinterpretation and misuse of results.

In time-sensitive contexts, such as humanitarian response, it may be appropriate to share timely but less precise estimates, provided that their limitations are clearly communicated. Transparency about uncertainty is a key component of responsible data use.

7.5 Managing Communications Within Partnerships

MPD initiatives are almost always collaborative, making internal communication and coordination within the partnership as important as external messaging.

Clear approval processes should be established for all partnership-related communications, including announcements, data releases, presentations, and reports. These processes should specify what requires approval, who provides it, how requests are made, and expected timelines.

Branding and recognition agreements are another important operational aspect. Partners should agree on how logos, names, and roles are presented, ensuring accurate representation and balanced visibility. Using agreed boilerplate language can help maintain consistency and avoid misunderstandings.

Above all, effective partnership communication relies on ongoing, two-way dialogue. Discussing communication principles early, revisiting them regularly, and seeking guidance when uncertainty arises helps protect both the initiative and the relationships that sustain it.

7.6 Conclusion

Managing communications in MPD initiatives is a strategic and operational responsibility, not a peripheral task. Clear, transparent, and audience-appropriate communication underpins trust, supports ethical practice, and enables data-driven insights to be used effectively for public benefit.

By applying general communication principles, addressing sensitivities specific to MPD, and embedding communication planning into project governance, organisations can ensure that such initiatives are not only technically sound, but also socially legitimate and institutionally sustainable.

A Appendix 1: Further recommended resources

This annotated bibliography provides verified references for the material covered in each chapter of this manual, together with additional resources for further reading. References marked with a chapter number in the main text are listed here in full. All URLs and DOIs were verified prior to publication. Practitioners are encouraged to check for updated editions before citing.

A.1 Chapter 1: Planning a Mobile Phone Data Initiative

UN Statistics Division (2019). Handbook on the Use of Mobile Phone Data for Official Statistics. United Nations. Available at: <https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf> → The foundational UN reference for MPD practitioners. Covers all key aspects of MPD planning, data pipeline design, quality assurance, and governance. Recommended as the primary companion text to this manual.

Blondel, V.D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10). DOI: 10.1140/epjds/s13688-015-0046-0. Available at: <https://link.springer.com/article/10.1140/epjds/s13688-015-0046-0> → A widely cited academic survey of CDR data analysis, covering social networks, mobility, urban planning, and development applications. Useful background reading for Sections 1.2.1 and 5.2.

Flowminder Foundation (2023). Flowminder standards in producing mobility and population estimates from call detail records in low- and middle-income countries. Flowminder Foundation. Available at: <https://www.flowminder.org/resources/publications-reports/flowminder-standards-in-producing-mobility-and-population-estimates-from-call-details-records-in-low-and-middle-income-countries> → Practical methodological standards from Flowminder covering CDR ingestion, cleaning, home/work detection, aggregation, bias adjustment, and quality assurance. Relevant to Chapters 1, 4, and 5.

GSMA (2016). Mobile Privacy Principles. GSMA. Available at: https://www.gsma.com/solutions-and-impact/connectivity-for-good/public-policy/gsma_resources/mobile-privacy-principles/ → The industry benchmark for privacy principles in mobile data use.

Additional resources: General introduction to Mobile Phone Data — UN CEBD Task Team website: <https://unstats.un.org/bigdata/task-teams/mobile-phone/> Maturity Assessment Framework tool: https://worldbank.github.io/GDF-MPD/docs/project-resources/maturity_assessment_framework.html Theory of Change guidance for MPD initiatives: <https://worldbank.github.io/GDF-MPD/docs/project-resources/theory-of-change.html>

A.2 Chapter 2: Policy Applications

UN-CEBD Task Team on Mobile Phone Data (n.d.). Methodological Guide on the Use of Mobile Phone Data: Dynamic Population Mapping. United Nations Statistics Division. Available at: <https://unstats.un.org/wiki/spaces/MPDDPM/overview> → Authoritative UN methodology for population mapping applications. See Section 2.3.1.

UN-CEBD Task Team on Mobile Phone Data (n.d.). Methodological Guide on the Use of Mobile Phone Data: Displacement and Disaster Statistics. United Nations Statistics Division. Available at: <https://unstats.un.org/wiki/spaces/MPDDS/overview> → Core reference for displacement tracking methodology. See Section 2.3.2.

UN-CEBD Task Team on Mobile Phone Data (n.d.). Methodological Guide on the Use of Mobile Phone Data: Tourism Statistics. United Nations Statistics Division. Available at: <https://unstats.un.org/wiki/display/MPDTS> → Full methodological guide covering concepts, methods, and case studies for tourism. See Section 2.3.4.

UN-CEBD Task Team on Mobile Phone Data (n.d.). Methodological Guide on the Use of Mobile Phone Data: Migration Statistics. United Nations Statistics Division. Available at: <https://unstats.un.org/wiki/spaces/MPDMS/overview> → Relevant to Section 2.3 and any references to migration applications.

Caceres, N., Wideberg, J.P., & Benitez, F.G. (2007). Deriving origin–destination data from a mobile phone network. *IET Intelligent Transport Systems*, 1(1), 15–26. DOI: 10.1049/iet-its:20060020. Available at: <https://digital-library.theiet.org/doi/10.1049/iet-its%3A20060020> → The foundational engineering paper demonstrating MPD-based origin–destination matrix derivation. See Section 2.3.5.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073–1076. DOI: 10.1126/science.aac4420. Available at: <https://www.science.org/doi/10.1126/science.aac4420> → The foundational Science paper on using CDR data to predict wealth distribution (Rwanda). See Section 2.4 and 2.4.1.

Aiken, E., Bedoya, G., Blumenstock, J., & Coville, A. (2022). Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603, 864–870. DOI: 10.1038/s41586-022-04484-9. Available at: <https://www.nature.com/articles/s41586-022-04484-9> → Peer-reviewed Nature paper documenting the Togo Novissi cash transfer targeting initiative. See Section 2.4.2.

Aiken, E. et al. (2022). Togo Novissi programme — World Bank Results Brief. World Bank. Available at: <https://www.worldbank.org/en/results/2021/04/13/prioritizing-the-poorest-and-most-vulnerable-in-west-africa-togo-s-novissi-platform-for-social-protection-uses-machine-l> → An accessible institutional reference on the Togo social protection programme for readers wanting a non-technical overview.

Additional resources: Dynamic Population Mapping (Netherlands): <https://www.cbs.nl/en-gb/background/2025/14/using-cell-phones-to-compute-dynamic-population-densities-safely> Disaster management and displacement case studies: <https://www.sciencedirect.com/science/article/> Migration statistics from MPD: <https://www.migrationdataportal.org/resource/exploring-use-mobile-phone-data-national-migration-statistics> Transport case studies (Latin America): <https://www.nommon.es/case-studies/monitoring-travel-demand-bogota-buenos-aires-world-bank-transit-insights/> Tourism statistics — Estonia: <https://positium.com/blog/estonia-leads-the-production-of-tourism-statistics-using-mobile-positioning-data> Poverty mapping with MPD: <https://www.flowminder.org/resources/publications-reports/mapping-poverty-using-mobile-phone-and-satellite-data>

A.3 Chapter 3: Arranging Partnerships and Data Access

Positium (2025). A Roadmap to Accessing Mobile Network Data for Statistics. Produced for the Global Partnership for Sustainable Development Data. Available at: <https://www.data4sdgs.org/roadmap-accessing-mobile-network-data-statistics> → An excellent step-by-step practical guide on accessing MNO data for statistical purposes. Recommended companion reading for Sections 3.1 and 3.8.

GSMA (2019). Big Data for Social Good: Mobile Network Operator Data Sharing. GSMA. Available at: https://www.gsma.com/solutions-and-impact/connectivity-for-good/external-affairs/wp-content/uploads/2019/09/Big-Data-AI-Ethics_web.pdf → GSMA guidance on data sharing models for social good. Useful background for Chapter 3 discussions on MNO incentives and data-sharing frameworks.

Additional resources: Accessing Mobile Network Operator data — webinar: <https://www.data4sdgs.org/accessing-mobile-data-national-strategies-and-challenges> MoU templates for MPD partnerships: https://worldbank.github.io/GDF-MPD/docs/project-resources/mou_templates.html

A.4 Chapter 4: Data Processing and Data Pipelines

UN Statistics Division (2019). Handbook on the Use of Mobile Phone Data for Official Statistics, Chapter 2: Data Sources, Attributes and General Data Extraction Process. United Nations. Available at: <https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf> → Chapter 2 of the UN Handbook is the primary reference for MPD data sources, pipeline structure, and extraction processes covered in this chapter.

Flowminder Foundation (2023). Flowminder standards in producing mobility and population estimates from call detail records in low- and middle-income countries. Flowminder Foundation. Available at: <https://www.flowminder.org/resources/publications-reports/flowminder-standards-in-producing-mobility-and-population-estimates-from-call-details-records-in-low-and-middle-income-countries> → Detailed methodological standards document covering CDR ingestion, cleaning, home/work detection, aggregation, bias adjustment, and quality assurance as described in this chapter.

Flowminder Foundation (2023). Correcting measurement biases in the detection of long and short stay locations in sparse Call Detail Records. Flowminder Foundation. Available at: <https://www.flowminder.org/resources/publications-reports/correcting-measurement-biases-in-the-detection-of-long-and-short-stay-locations-in-sparse-call-detail-records-cdrs> → Technical paper on bias correction relevant to Sections 4.7 and 5.7 (scaling and bias adjustment).

Blondel, V.D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10). DOI: 10.1140/epjds/s13688-015-0046-0. Available at: <https://link.springer.com/article/10.1140/epjds/s13688-015-0046-0> → Broad survey of CDR data analysis methods including pipeline approaches relevant to Chapter 4.

A.5 Chapter 5: Data Quality and Characteristics

Flowminder Foundation (2023). Flowminder standards in producing mobility and population estimates from call detail records in low- and middle-income countries. Flowminder Foundation. Available at: <https://www.flowminder.org/resources/publications-reports/flowminder-standards-in-producing-mobility-and-population-estimates-from-call-details-records-in-low-and-middle-income-countries> → Covers CDR strengths and limitations, spatial and temporal precision, and methodological standards for quality assurance, all directly relevant to Chapter 5.

Flowminder Foundation (2023). Correcting measurement biases in the detection of long and short stay locations in sparse Call Detail Records. Flowminder Foundation. Available at: <https://www.flowminder.org/resources/publications-reports/correcting-measurement-biases-in-the-detection-of-long-and-short-stay-locations-in-sparse-call-detail-records-cdrs> → Peer-reviewed technical paper on bias correction in CDR data. Directly relevant to Section 5.7 on adjusting for bias.

Blondel, V.D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10). DOI: 10.1140/epjds/s13688-015-0046-0. Available at: <https://link.springer.com/article/10.1140/epjds/s13688-015-0046-0> → See for further details relevant to the overview of CDR strengths and limitations in Section 5.2.

Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., & Buckee, C.O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface*, 10(81), 20120986. DOI: 10.1098/rsif.2012.0986. → The foundational paper on representativeness bias in CDR data, covering differential phone ownership by gender, age, and wealth. Essential citation for Section 5.6. Note: this paper is paywalled and must be accessed via institutional library subscription.

A.6 Chapter 6: Data Governance and Safeguards

GSMA (2016). *Mobile Privacy Principles*. GSMA. Available at: https://www.gsma.com/solutions-and-impact/connectivity-for-good/public-policy/gsma_resources/mobile-privacy-principles/ → The industry benchmark for privacy principles in mobile data use. Cited in Sections 1.4 and 6.8.3.

Jansen, R. et al. (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. → The peer-reviewed paper behind the UN guiding principles described in Section 6.8.1. Provides the full evidence base and rationale for the five principles.

United Nations (2014). *Fundamental Principles of Official Statistics*. United Nations Statistics Division. Available at: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx> → The UN Fundamental Principles apply to data governance, quality assurance, and public trust in official statistics. See Sections 6.2, 6.8, and throughout Chapter 6 where governance obligations are discussed.

African Union (2018). *African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention)*. African Union. Available at: <https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection> → The continental African framework for data protection.

References

- African Union. 2018. *African Union Convention on Cyber Security and Personal Data Protection*. <https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection>.
- Ahas, Rein, Anto Aasa, Antti Roose, Ülar Mark, and Siiri Silm. 2008. "Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study." *Tourism Management* 29 (3): 469–86. <https://doi.org/10.1016/j.tourman.2007.05.014>.
- Aiken, Emily, Guadalupe Bedoya, Joshua Blumenstock, and Aidan Coville. 2022. "Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid." *Nature* 603: 864–70. <https://doi.org/10.1038/s41586-022-04484-9>.
- Alexander, Lauren, Shan Jiang, Mikel Murga, and Marta C. Gonzalez. 2015. "Origin-Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data." *Transportation Research Part C: Emerging Technologies* 58: 240–50. <https://doi.org/10.1016/j.trc.2015.02.018>.
- Ascari, G., E. Cerasti, C. Faricelli, et al. 2024. "Quality Aspects Using Mobile Network Operators Data for Official Statistics." *2nd Workshop on Methodologies for Official Statistics: Proceedings* (Rome), 135–51. <https://www.istat.it/wp-content/uploads/2024/11/2nd-Workshop-on-methodologies-for-official-statistics-Proceedings-1.pdf>.
- Barreras, Francisco, and Duncan J. Watts. 2024. "The Exciting Potential and Daunting Challenge of Using GPS Human-Mobility Data for Epidemic Modeling." *Nature Computational Science* 4 (6): 398–411. <https://doi.org/10.1038/s43588-024-00637-0>.
- Benchmark Initiative. 2021. *The Locus Charter*. <https://ethicalgeo.org/locus-charter/>.
- Bengtsson, Linus, Jean Gaudart, Xin Lu, et al. 2015. "Using Mobile Phone Data to Predict the Spatial Spread of Cholera." *Scientific Reports* 5: 8923. <https://doi.org/10.1038/srep08923>.
- Blondel, Vincent D., Adeline Decuyper, and Gautier Krings. 2015. "A Survey of Results on Mobile Phone Datasets Analysis." *EPJ Data Science* 4 (10). <https://doi.org/10.1140/epjds/s13688-015-0046-0>.

- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76. <https://doi.org/10.1126/science.aac4420>.
- Cabrera, Carmen et al. 2025. *Sustained Changes to Urban Mobility After COVID-19 Amplified Socio-Economic Inequalities in Latin America*. <https://doi.org/10.48550/arXiv.2504.15871>.
- Cabrera, Carmen, and Francisco Rowe. 2025. *A Systematic Machine Learning Approach to Measure and Assess Biases in Mobile Phone Population Data*. <https://doi.org/10.48550/arXiv.2509.02603>.
- Caceres, N., J. P. Wideberg, and F. G. Benitez. 2007. "Deriving Origin-Destination Data from a Mobile Phone Network." *IET Intelligent Transport Systems* 1 (1): 15–26. <https://doi.org/10.1049/iet-its:20060020>.
- Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. "Estimating Origin-Destination Flows Using Mobile Phone Location Data." *IEEE Pervasive Computing* 10 (4): 36–44. <https://doi.org/10.1109/MPRV.2011.41>.
- Deville, Pierre, Catherine Linard, Samuel Martin, et al. 2014. "Dynamic Population Mapping Using Mobile Phone Data." *Proceedings of the National Academy of Sciences* 111 (45): 15888–93. <https://doi.org/10.1073/pnas.1408439111>.
- Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network Diversity and Economic Development." *Science* 328 (5981): 1029–31. <https://doi.org/10.1126/science.1186605>.
- European Statistical System. 2019. *Quality Assurance Framework of the European Statistical System, Version 2.0*. <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>.
- FlowGeek. n.d. *Data Governance and Data Privacy for Mobile Phone Data*. <https://flowgeek.org/>.
- Flowminder Foundation. 2023a. *Correcting Measurement Biases in the Detection of Long and Short Stay Locations in Sparse Call Detail Records*. <https://www.flowminder.org/resources/publications-reports/correcting-measurement-biases-in-the-detection-of-long-and-short-stay-locations-in-sparse-call-detail-records-cdrs>.
- Flowminder Foundation. 2023b. *Flowminder Standards in Producing Mobility and Population Estimates from Call Detail Records in Low- and Middle-Income Countries*. <https://www.flowminder.org/resources/publications-reports/flowminder-standards-in-producing-mobility-and-population-estimates-from-call-details-records-in-low-and-middle-income-countries>.
- GDPR.eu. n.d. *General Data Protection Regulation Compliance Overview*. <https://gdpr.eu/>.

- Global Partnership for Sustainable Development Data, and Positium. 2025. *A Roadmap to Accessing Mobile Network Data for Statistics*. <https://www.data4sdgs.org/roadmap-accessing-mobile-network-data-statistics>.
- Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. 2008. "Understanding Individual Human Mobility Patterns." *Nature* 453 (7196): 779–82. <https://doi.org/10.1038/nature06958>.
- GSMA. 2016. *Mobile Privacy Principles*. https://www.gsma.com/solutions-and-impact/connectivity-for-good/public-policy/gsma_resources/mobile-privacy-principles/.
- GSMA. 2019. *Big Data for Social Good: Mobile Network Operator Data Sharing*. https://www.gsma.com/solutions-and-impact/connectivity-for-good/external-affairs/wp-content/uploads/2019/09/Big-Data-AI-Ethics_web.pdf.
- International Telecommunication Union. n.d.a. *Big Data for Measuring the Information Society*. https://www.itu.int/en/ITU-D/Statistics/Documents/bigdata/ITU_SDG_case_study.pdf.
- International Telecommunication Union. n.d.b. *ITU Academy*. <https://academy.itu.int/>.
- Iradukunda, Rodgers, Francisco Rowe, and Elisabetta Pietrostefani. 2025. "Estimating Internal Displacement in Ukraine from High-Frequency GPS Mobile Phone Data." *Humanities and Social Sciences Communications*, ahead of print. <https://doi.org/10.1057/s41599-025-06137-4>.
- Isaacman, Sibren, Richard Becker, Ramon Caceres, et al. 2011. "Identifying Important Places in People's Lives from Cellular Network Data." *Pervasive Computing*, Lecture notes in computer science, vol. 6696: 133–51. https://doi.org/10.1007/978-3-642-21726-5_9.
- Jansen, Ronald, Miglena Abels, Savvas Papadakis, Benjamin Sakarovitch, et al. 2021. "Guiding Principles to Maintain Public Trust in the Use of Mobile Operator Data for Policy Purposes." *Data & Policy* 3: E24. <https://doi.org/10.1017/dap.2021.21>.
- Li, Tong, Cordelia Bowers, Alhassan Seidu, Gloria Akoto-Bamfo, Daniel Ofori-Boateng, et al. 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>.
- Louail, Thomas, Maxime Lenormand, Oliva G. Cantu Ros, et al. 2014. "From Mobile Phone Data to the Spatial Structure of Cities." *Scientific Reports* 4: 5276. <https://doi.org/10.1038/srep05276>.

- Lu, Xin, Linus Bengtsson, and Petter Holme. 2012. "Predictability of Population Displacement After the 2010 Haiti Earthquake." *Proceedings of the National Academy of Sciences* 109 (29): 11576–81. <https://doi.org/10.1073/pnas.1203882109>.
- Montjoye, Yves-Alexandre de, Sebastien Gambs, Vincent D. Blondel, et al. 2018. "On the Privacy-Conscientious Use of Mobile Phone Data." *Scientific Data* 5: 180286. <https://doi.org/10.1038/sdata.2018.286>.
- Montjoye, Yves-Alexandre de, Cesar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3: 1376. <https://doi.org/10.1038/srep01376>.
- Pietro Stefani, Elisabetta, Matt Mason, Rodgers Iradukunda, Hong Tran-Jones, Iryna Loktiewa, and Francisco Rowe. 2025. *Dynamic Estimates of Displacement in Disaster Regions: A Policy-Driven Framework Triangulating Data*. <https://doi.org/10.48550/arXiv.2511.01955>.
- Ricciato, Fabio, Giampaolo Lanzieri, Albrecht Wirthmann, and Gerdy Seynaeve. 2020. "Towards a Methodological Framework for Estimating Present Population Density from Mobile Network Operator Data." *Pervasive and Mobile Computing* 68: 101263. <https://doi.org/10.1016/j.pmcj.2020.101263>.
- Rowe, F., E. Magpantay, M. Jalagonia, et al. 2022. *Methodological Guide on the Use of Mobile Phone Data: Migration Statistics*. United Nations Statistics Division; UN-CEBD Task Team on Mobile Phone Data. <https://unstats.un.org/wiki/spaces/MPDMS/overview>.
- Rowe, Francisco. 2022. "Using Digital Footprint Data to Monitor Human Mobility and Support Rapid Humanitarian Responses." *Regional Studies, Regional Science*, ahead of print. <https://doi.org/10.1080/21681376.2022.2135458>.
- Rowe, Francisco, Alessia Calafiore, Daniel Arribas-Bel, Konstantin Samardzhiev, and Martin Fleischmann. 2023. "Urban Exodus? Understanding Human Mobility in Britain During the COVID-19 Pandemic Using Meta-Facebook Data." *Population, Space and Place*, ahead of print. <https://doi.org/10.1002/psp.2637>.
- Rowe, Francisco, and Miguel González-Leonardo. 2024. *Digital Data and Population Studies*. <http://dx.doi.org/10.31219/osf.io/jb3e5>.
- Salgado, David, Luis Sanguiao, Bogdan Oancea, Sandra Barragan, and Marian Necula. 2021. "An End-to-End Statistical Process with Mobile Network Data for Official Statistics." *EPJ Data Science* 10 (20). <https://doi.org/10.1140/epjds/s13688-021-00275-w>.
- Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi. 2010. "Limits of Predictability in Human Mobility." *Science* 327 (5968): 1018–21. <https://doi.org/10.1126/science.1177170>.

- Steele, Jessica E., Pal Roe Sundsoy, Carla Pezzulo, et al. 2017. "Mapping Poverty Using Mobile Phone and Satellite Data." *Journal of the Royal Society Interface* 14 (127): 20160690. <https://doi.org/10.1098/rsif.2016.0690>.
- Tizzoni, Michele, Paolo Bajardi, Adeline Decuyper, et al. 2014. "On the Use of Human Mobility Proxies for Modeling Epidemics." *PLOS Computational Biology* 10 (7): e1003716. <https://doi.org/10.1371/journal.pcbi.1003716>.
- Toole, Jameson L., Serdar Colak, Bradley Sturt, Lauren P. Alexander, Alexandre Evsukoff, and Marta C. Gonzalez. 2015. "The Path Most Traveled: Travel Demand Estimation Using Big Data Resources." *Transportation Research Part C: Emerging Technologies* 58: 162–77. <https://doi.org/10.1016/j.trc.2015.04.022>.
- UN-CEBD Task Team on Mobile Phone Data. n.d.a. *Methodological Guide on the Use of Mobile Phone Data: Displacement and Disaster Statistics*. United Nations Statistics Division. <https://unstats.un.org/wiki/spaces/MPDDS/overview>.
- UN-CEBD Task Team on Mobile Phone Data. n.d.b. *Methodological Guide on the Use of Mobile Phone Data: Dynamic Population Mapping*. United Nations Statistics Division. <https://unstats.un.org/wiki/spaces/MPDDPM/overview>.
- UN-CEBD Task Team on Mobile Phone Data. n.d.c. *Methodological Guide on the Use of Mobile Phone Data: Measuring the Information Society*. United Nations Statistics Division. <https://unstats.un.org/wiki/spaces/MPDMIS/overview>.
- UN-CEBD Task Team on Mobile Phone Data. n.d.d. *Methodological Guide on the Use of Mobile Phone Data: Tourism Statistics*. United Nations Statistics Division. <https://unstats.un.org/wiki/display/MPDTS>.
- United Nations. 2014. *Fundamental Principles of Official Statistics*. <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.
- United Nations Committee of Experts on Big Data and Data Science for Official Statistics. n.d. *Task Team on Mobile Phone Data*. <https://unstats.un.org/bigdata/task-teams/mobile-phone/>.
- United Nations Statistics Division. 2019. *Handbook on the Use of Mobile Phone Data for Official Statistics*. United Nations. <https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf>.
- Wesolowski, Amy, Nathan Eagle, Andrew J. Tatem, et al. 2012. "Quantifying the Impact of Human Mobility on Malaria." *Science* 338 (6104): 267–70. <https://doi.org/10.1126/science.1223467>.

Wesolowski, Amy, Nathan Eagle, Andrew J. Tatem, et al. 2013. "The Impact of Biases in Mobile Phone Ownership on Estimates of Human Mobility." *Journal of the Royal Society Interface* 10 (81): 20120986. <https://doi.org/10.1098/rsif.2012.0986>.

World Bank. 2021. *Togo's Novissi Platform for Social Protection Uses Machine Learning, Mobile Money, and Satellite Data to Support the Most Vulnerable*. <https://www.worldbank.org/en/results/2021/04/13/prioritizing-the-poorest-and-most-vulnerable-in-west-africa-togo-s-novissi-platform-for-social-protection-uses-machine-l>.