ITU WTIS

GENEVA 2023

# Preparing the ground: Mobile phone data for official statistics

Applications, guides and way forward

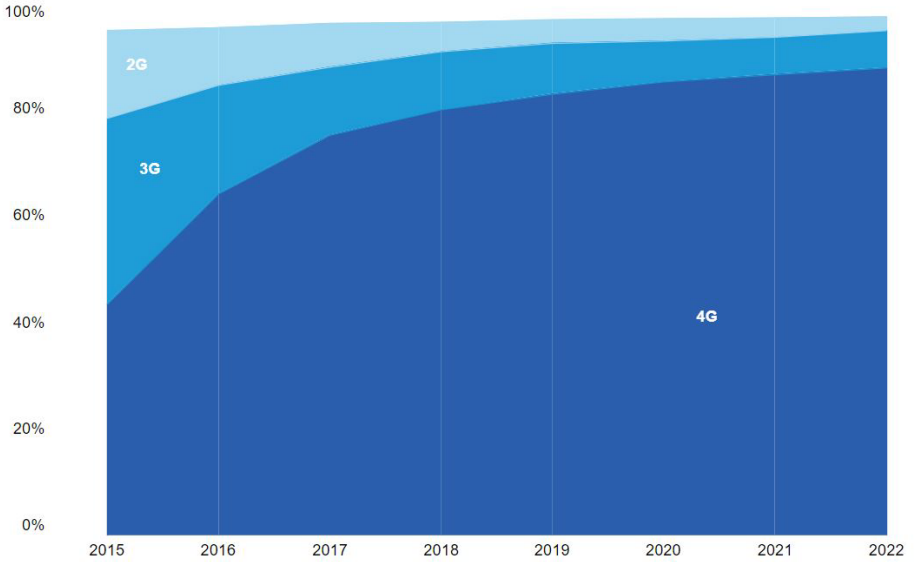
4 July 2023 | 9:30-11:00 (CET)

ITU WTIS GENEVA2023

Esperanza Magpantay
Senior Statistician, ICT Data and Analytics division, BDT/ITU
Chair, UN-CEBD Task Team on Mobile phone data

# Two-thirds of the world's population use the Internet, but 2.7 billion people remain offline

Population coverage by type of mobile network, 2015-2022

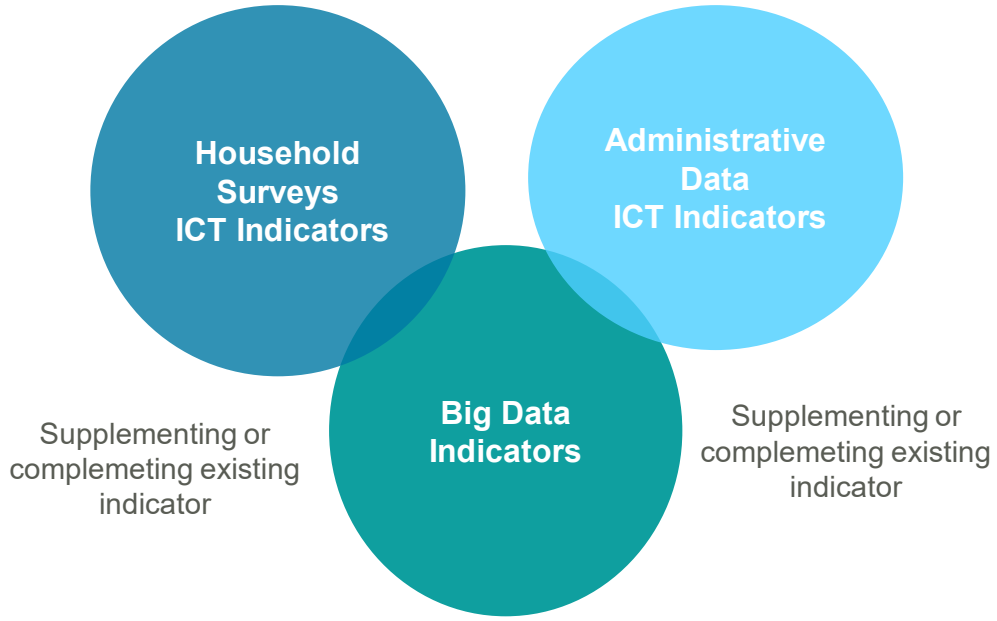



**Note:** The values for 2G and 3G networks show the incremental percentage of the population that is not covered by a more advanced technology network (e.g. in 2022, 95 per cent of the world population is covered by a 3G or above network, that is 7 per cent + 88 per cent).

Source: ITU


Countries with Internet use data (2019-2021) – ~110 countries

## Background: Big Data for Measuring the Information Society




### Pilots: 2016 – 2020

- Brazil
- Colombia
- Georgia
- Indonesia
- Kenya
- Philippines
- United Arab Emirates

### Stakeholders

- Ministry
- Regulator
- Data protection agency
- Operators
- National Statistics Office

### Challenges:

- Data access
- Lack of capacity/data scientists
- Lack of IT resources

- Big data is revolutionizing the world of statistics.
- Huge opportunities: more granular, more timely, more accurate, less costly to collect, more insightful, more valuable.
- Since 2016, ITU implemented mobile phone big data projects in several countries.
- Along the way, we have refined our methods and models and created guidelines for countries exploring big data.
- Coordination among different stakeholders is key!



# UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)

## Task Teams:

- **Mobile phone data**
- Earth observation data
- Scanner data
- AIS data
- Privacy preserving techniques
- Access to privately-held data
- Big data for SDGs
- Training and Capacity Building

## Regional Hubs

- Brazil
- China
- Rwanda
- UAE

UNBigData

# UN-CEBD Task Team on mobile phone data

UNBigData

## Chair: ITU

### Objectives

Explore the use of mobile phone big data for the different areas of statistics and develop methodologies

### Who

Composed of around 50 individual members/ 30 entities - international and regional agencies, countries, academia, private agencies/companies

### Members

- Brazil
- Colombia
- Gambia
- Georgia
- India
- Indonesia
- Italy
- Japan
- Korea
- Malaysia

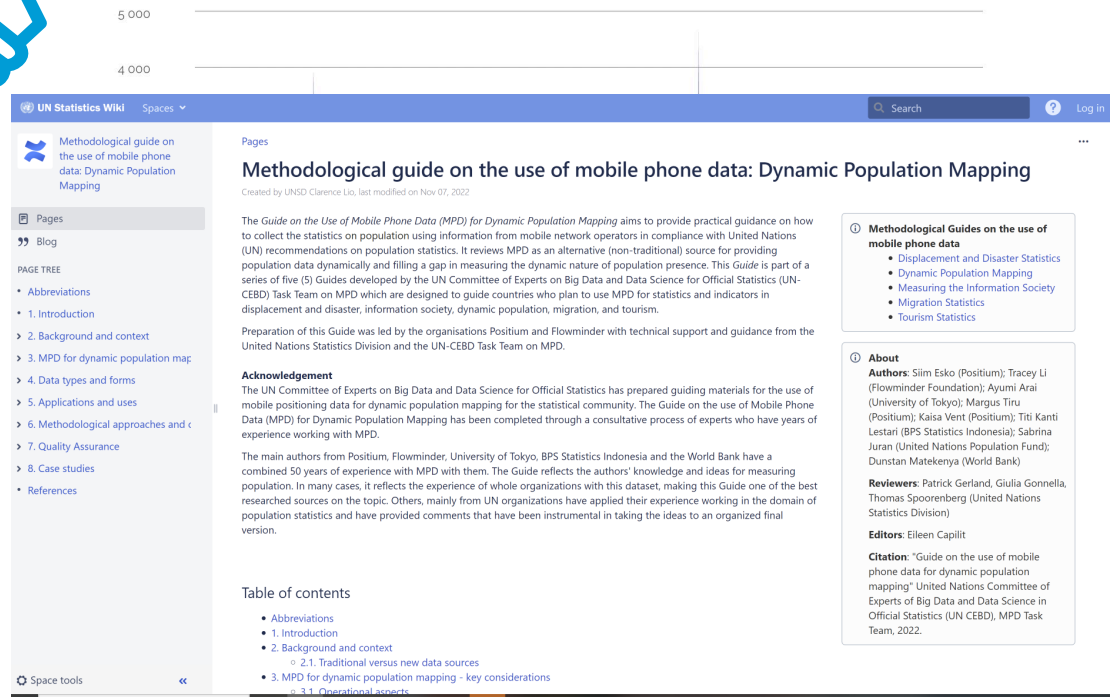
### Members

- Mexico
- Netherlands
- Oman
- Qatar
- Philippines
- Romania
- Saudi Arabia
- United Arab Emirates

### Members

- EU JRC
- Eurostat
- IMF
- IOM
- UNFPA
- UNGP Jakarta
- UNSD
- UNESCWA
- World Bank
- OECD-ITF
- UN-ECE
- Flowminder
- GSMA
- Positium
- Telenor

# Dynamic Population Mapping




Aims to answer - how to **map population dynamically** - without being dependent on logistics of surveys or the census

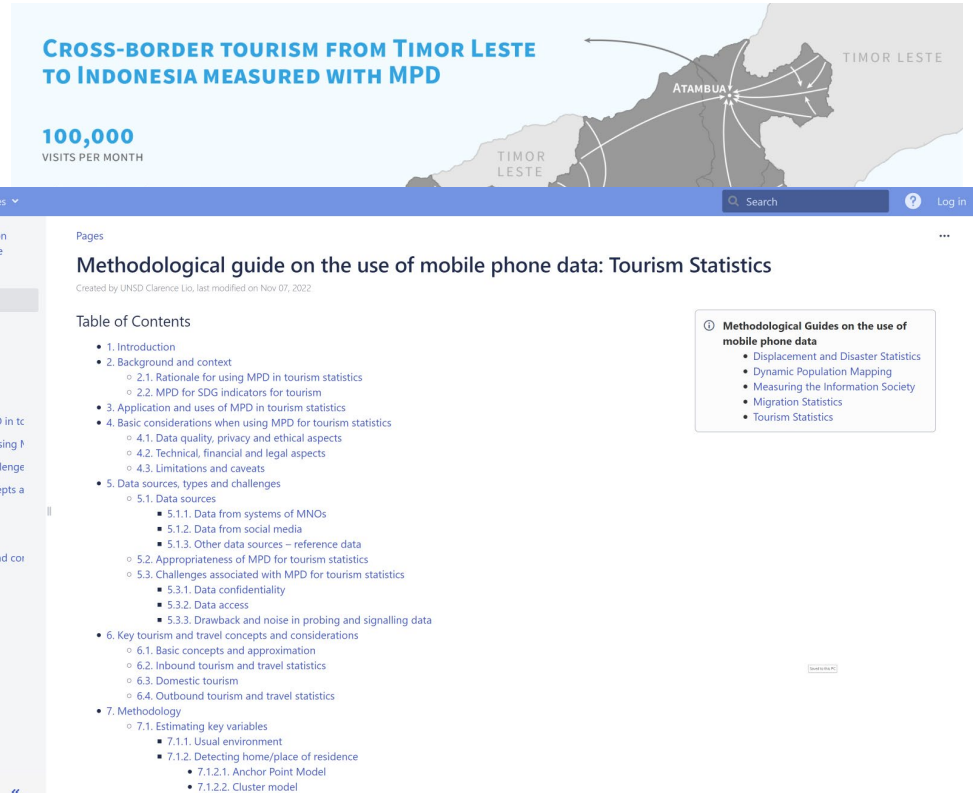
& how to **map a dynamically-behaving population** - understanding de facto presence at any time, even away from place of residence

MPD use cases:

1. Resident population mapping
2. Daytime population mapping
3. De facto population mapping
4. Monitoring population redistributions caused by COVID-19 mobility restrictions
5. Infrastructure and resource planning
6. Creating dynamic sample frames for surveys
7. Census
8. Disaster preparedness planning and response

See: Methodological guide on the use of mobile phone data: Dynamic population mapping, https://unstats.un.org/wiki/display/MPDDPM

# Tourism statistics




Challenges:

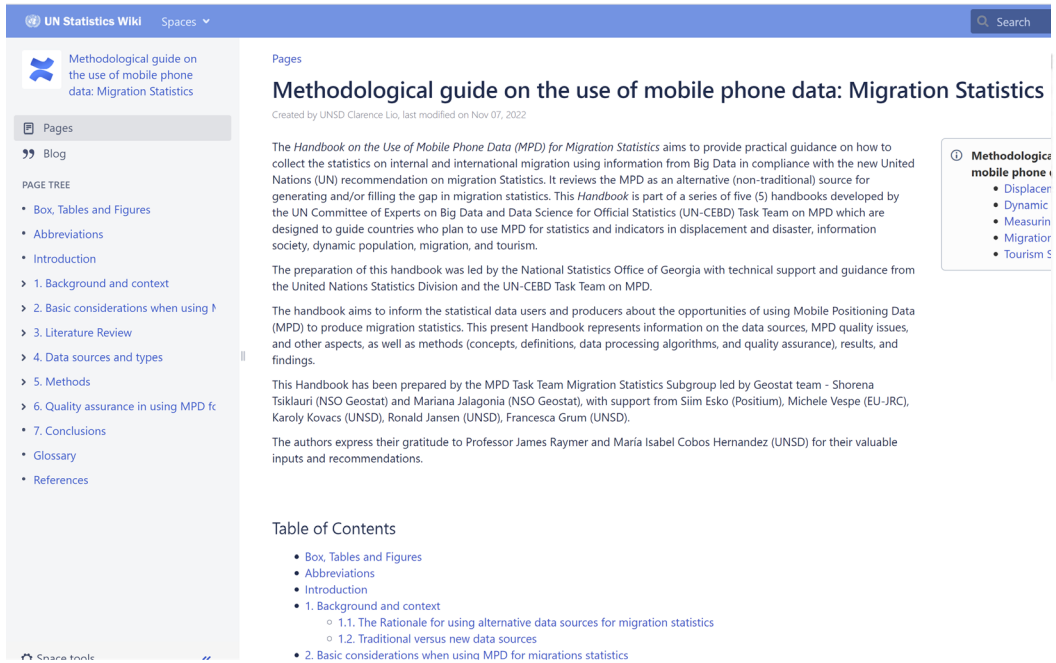
- Existing data sources had coverage issues.
- Surveys were limited in scope, with data collection only taking place over a month and only in select locations to estimate results for one-year time period for the whole border.
- Immigration data did not account for areas where there is no border checkpoint
- Conducting surveys was expensive

Implementation:

- BPS-Statistics Indonesia has been using MPD for official tourism statistics since 2016
- Implemented to measure mobilities within Indonesia's border regions - movements to and from the country

See: Methodological guide on the use of mobile phone data: Tourism Statistics, https://unstats.un.org/wiki/display/MPDTS

# Migration statistics



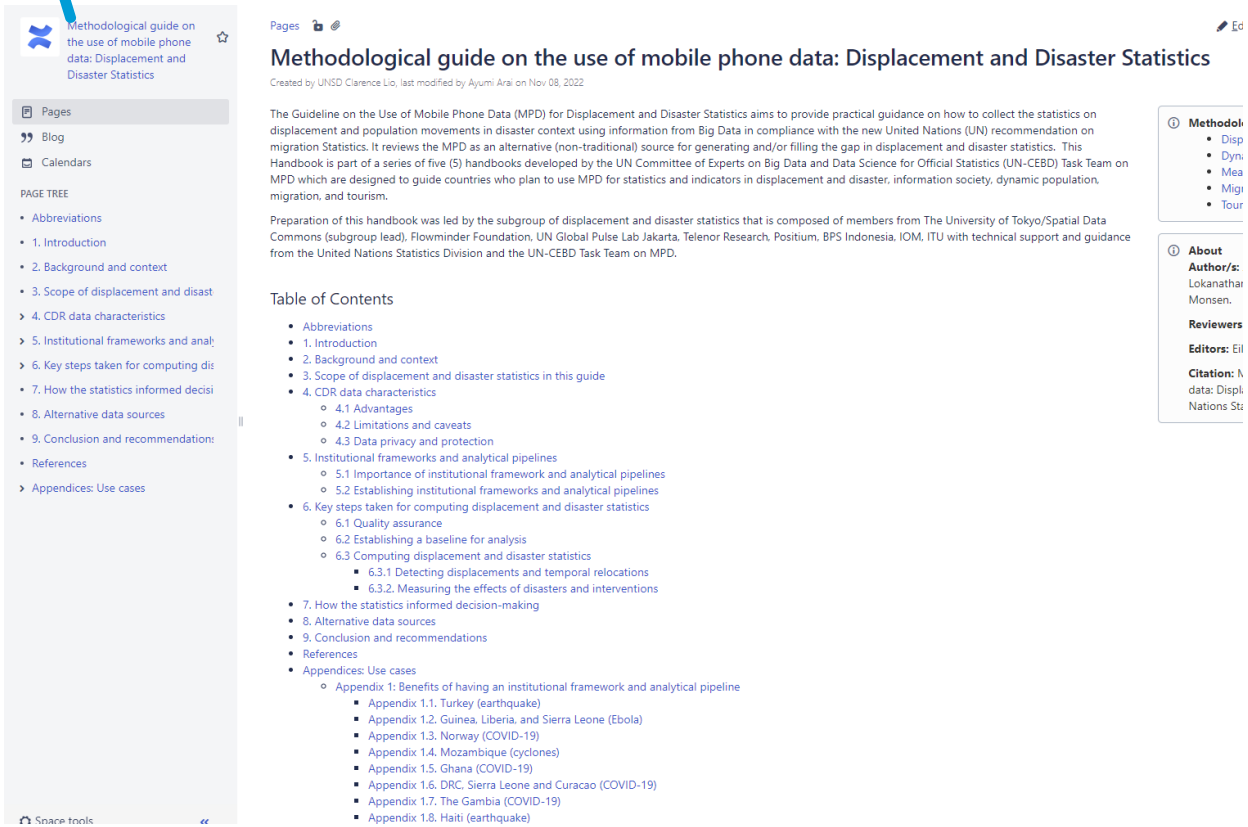

The use of MPD can improve several aspects of migration statistics:

- timeliness
- access to statistical information previously unavailable
- calibration opportunities for existing data
- granularity

*Comparison of flows for internal migration between counties in Estonia according to census 2011 and MPD during the period 2010–2011. N= number of migrants between counties (LAU level 1).*

See: Methodological guide on the use of mobile phone data: migration statistics, https://unstats.un.org/wiki/display/MPDMS

# Displacement and Disaster statistics



Methodological guide on the use of mobile phone data: Displacement and Disaster Statistics

Pages
Blog
Calendars

PAGE TREE
- Abbreviations
- 1. Introduction
- 2. Background and context
- 3. Scope of displacement and disast
- 4. CDR data characteristics
- 5. Institutional frameworks and anal
- 6. Key steps taken for computing dis
- 7. How the statistics informed decisi
- 8. Alternative data sources
- 9. Conclusion and recommendations
- References
- Appendices: Use cases

Space tools

Pages

## Methodological guide on the use of mobile phone data: Displacement and Disaster Statistics

Created by UNSD Clarence Lio, last modified by Ayumi Arai on Nov 08, 2022

The Guideline on the Use of Mobile Phone Data (MPD) for Displacement and Disaster Statistics aims to provide practical guidance on how to collect the statistics on displacement and population movements in disaster context using information from Big Data in compliance with the new United Nations (UN) recommendation on migration Statistics. It reviews the MPD as an alternative (non-traditional) source for generating and/or filling the gap in displacement and disaster statistics. This Handbook is part of a series of five (5) handbooks developed by the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) Task Team on MPD which are designed to guide countries who plan to use MPD for statistics and indicators in displacement and disaster, information society, dynamic population, migration, and tourism.

Preparation of this handbook was led by the subgroup of displacement and disaster statistics that is composed of members from The University of Tokyo/Spatial Data Commons (subgroup lead), Flowminder Foundation, UN Global Pulse Lab Jakarta, Telenor Research, Positium, BPS Indonesia, IOM, ITU with technical support and guidance from the United Nations Statistics Division and the UN-CEBD Task Team on MPD.

Table of Contents

- Abbreviations
- 1. Introduction
- 2. Background and context
- 3. Scope of displacement and disaster statistics in this guide
- 4. CDR data characteristics
  - 4.1 Advantages
  - 4.2 Limitations and caveats
  - 4.3 Data privacy and protection
- 5. Institutional frameworks and analytical pipelines
  - 5.1 Importance of institutional framework and analytical pipelines
  - 5.2 Establishing institutional frameworks and analytical pipelines
- 6. Key steps taken for computing displacement and disaster statistics
  - 6.1 Quality assurance
  - 6.2 Establishing a baseline for analysis
  - 6.3 Computing displacement and disaster statistics
    - 6.3.1 Detecting displacements and temporal relocations
    - 6.3.2. Measuring the effects of disasters and interventions
- 7. How the statistics informed decision-making
- 8. Alternative data sources
- 9. Conclusion and recommendations
- References
- Appendices: Use cases
  - Appendix 1: Benefits of having an institutional framework and analytical pipeline
    - Appendix 1.1. Turkey (earthquake)
    - Appendix 1.2. Guinea, Liberia, and Sierra Leone (Ebola)
    - Appendix 1.3. Norway (COVID-19)
    - Appendix 1.4. Mozambique (cyclones)
    - Appendix 1.5. Ghana (COVID-19)
    - Appendix 1.6. DRC, Sierra Leone and Curacao (COVID-19)
    - Appendix 1.7. The Gambia (COVID-19)
    - Appendix 1.8. Haiti (earthquake)

Possible to estimate:

- the numbers of people displaced from the areas most directly impacted by the earthquake
- the areas people had been displaced to
- And the disruption to transportation in the affected areas

See: Methodological guide on the use of mobile phone data: displacement and disaster statistics, https://unstats.un.org/wiki/display/MPDDS/

## Transport statistics

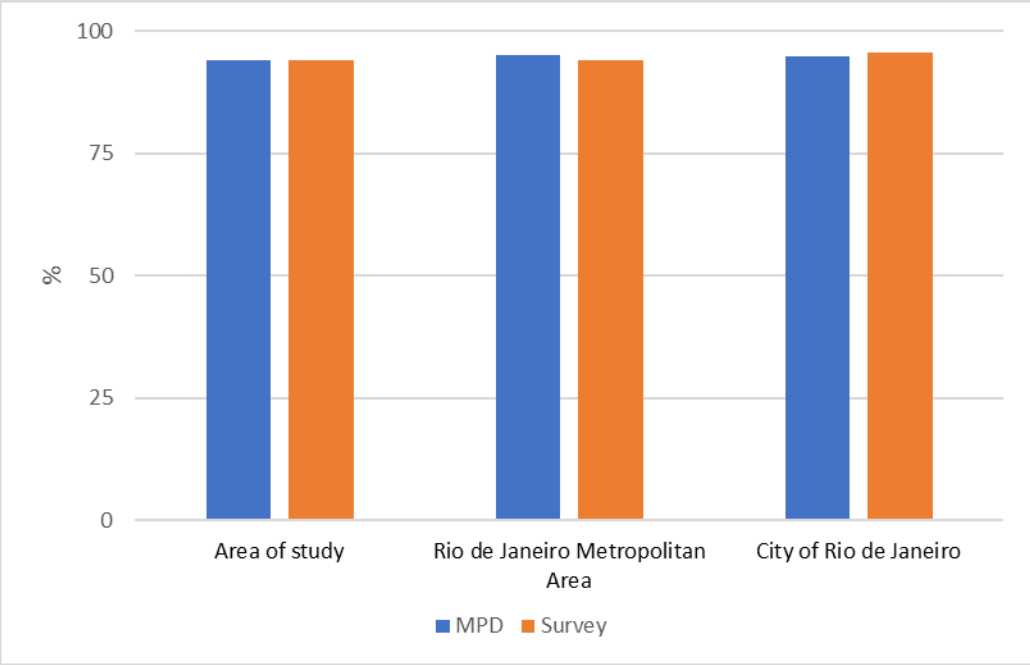
- ✓Origin-Destination matrices between regions
- ✓Public transport statistics (particularly when monthly passes are used)
- ✓Urban mobility, especially walking and cycling
- ✓Transport planning

See: Methodological guide on the use of mobile phone data: Transport Statistics (to be available soon)

UNBigData

# Information society (SDG) indicators

Percentage of the population using the Internet, Brazil, 2021

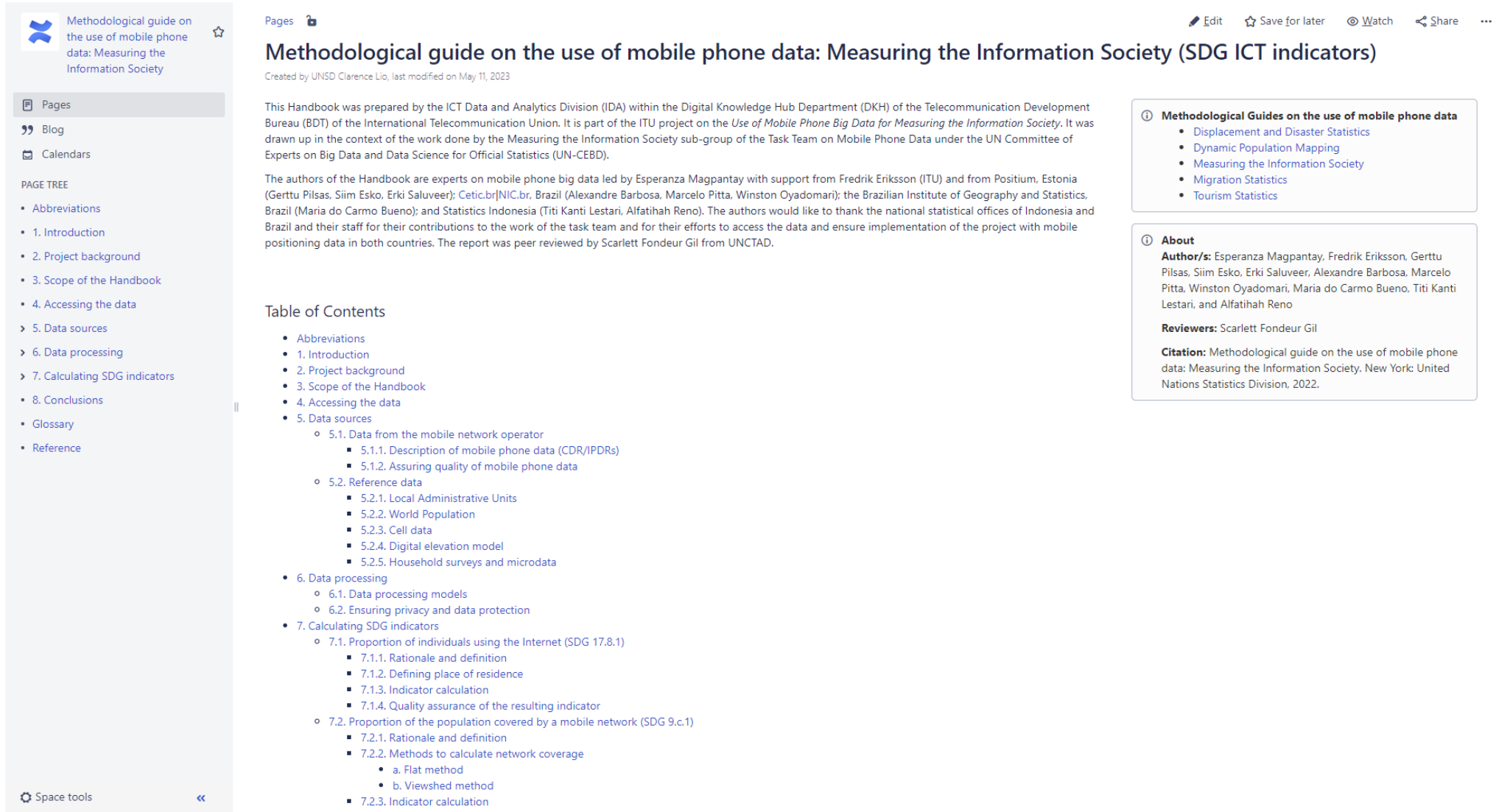



- Surveys to collect data on Internet use are not conducted in many countries - lack of resources
- The lack of official data presents a challenge for monitoring SDGs

- ✓ suggest that MPD can be used to calculate the two SDG indicators - Brazil and Indonesia
- ✓ The two SDG ICT indicators - timely and with greater spatial resolution

See: Methodological guide on the use of mobile phone data: information society SDG indicators, https://unstats.un.org/wiki/display/MPDMIS

# Methodological guide on the use of mobile phone data: measuring the information society indicators



Methodological guide on the use of mobile phone data: Measuring the Information Society

Pages
Blog
Calendars

PAGE TREE

- Abbreviations
- 1. Introduction
- 2. Project background
- 3. Scope of the Handbook
- 4. Accessing the data
- 5. Data sources
- 6. Data processing
- 7. Calculating SDG indicators
- 8. Conclusions
- Glossary
- Reference

Space tools

Pages

Edit | Save for later | Watch | Share

## Methodological guide on the use of mobile phone data: Measuring the Information Society (SDG ICT indicators)

Created by UNSD Clarence Lio, last modified on May 11, 2023

This Handbook was prepared by the ICT Data and Analytics Division (IDA) within the Digital Knowledge Hub Department (DKH) of the Telecommunication Development Bureau (BDT) of the International Telecommunication Union. It is part of the ITU project on the *Use of Mobile Phone Big Data for Measuring the Information Society*. It was drawn up in the context of the work done by the Measuring the Information Society sub-group of the Task Team on Mobile Phone Data under the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD).

The authors of the Handbook are experts on mobile phone big data led by Esperanza Magpantay with support from Fredrik Eriksson (ITU) and from Positium, Estonia (Gerttu Pilsas, Siim Esko, Erki Saluveer); Cetic.br|NIC.br, Brazil (Alexandre Barbosa, Marcelo Pitta, Winston Oyadomari); the Brazilian Institute of Geography and Statistics, Brazil (Maria do Carmo Bueno); and Statistics Indonesia (Titi Kanti Lestari, Alfatihah Reno). The authors would like to thank the national statistical offices of Indonesia and Brazil and their staff for their contributions to the work of the task team and for their efforts to access the data and ensure implementation of the project with mobile positioning data in both countries. The report was peer reviewed by Scarlett Fondeur Gil from UNCTAD.

**Methodological Guides on the use of mobile phone data**

- Displacement and Disaster Statistics
- Dynamic Population Mapping
- Measuring the Information Society
- Migration Statistics
- Tourism Statistics

**About**

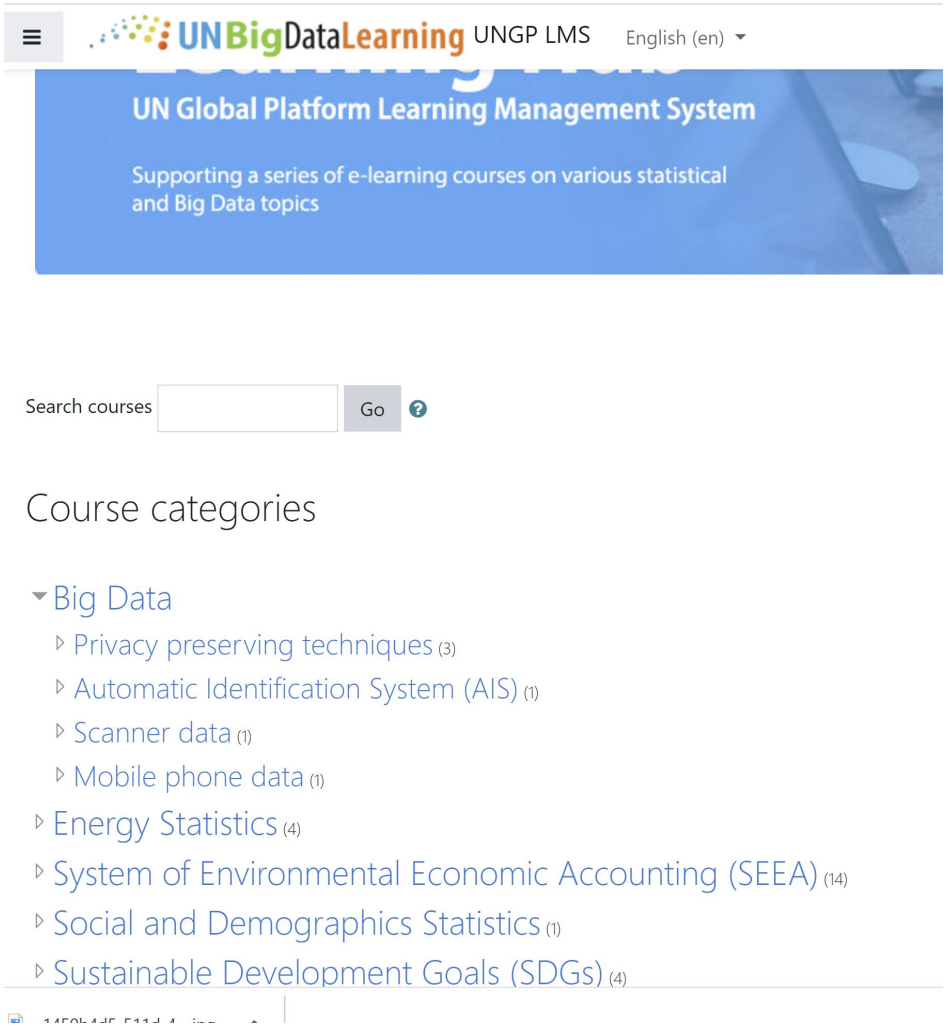
**Author/s:** Esperanza Magpantay, Fredrik Eriksson, Gerttu Pilsas, Siim Esko, Erki Saluveer, Alexandre Barbosa, Marcelo Pitta, Winston Oyadomari, Maria do Carmo Bueno, Titi Kanti Lestari, and Alfatihah Reno

**Reviewers:** Scarlett Fondeur Gil

**Citation:** Methodological guide on the use of mobile phone data: Measuring the Information Society. New York: United Nations Statistics Division, 2022.

Table of Contents

# Mobile phone data – awareness raising course



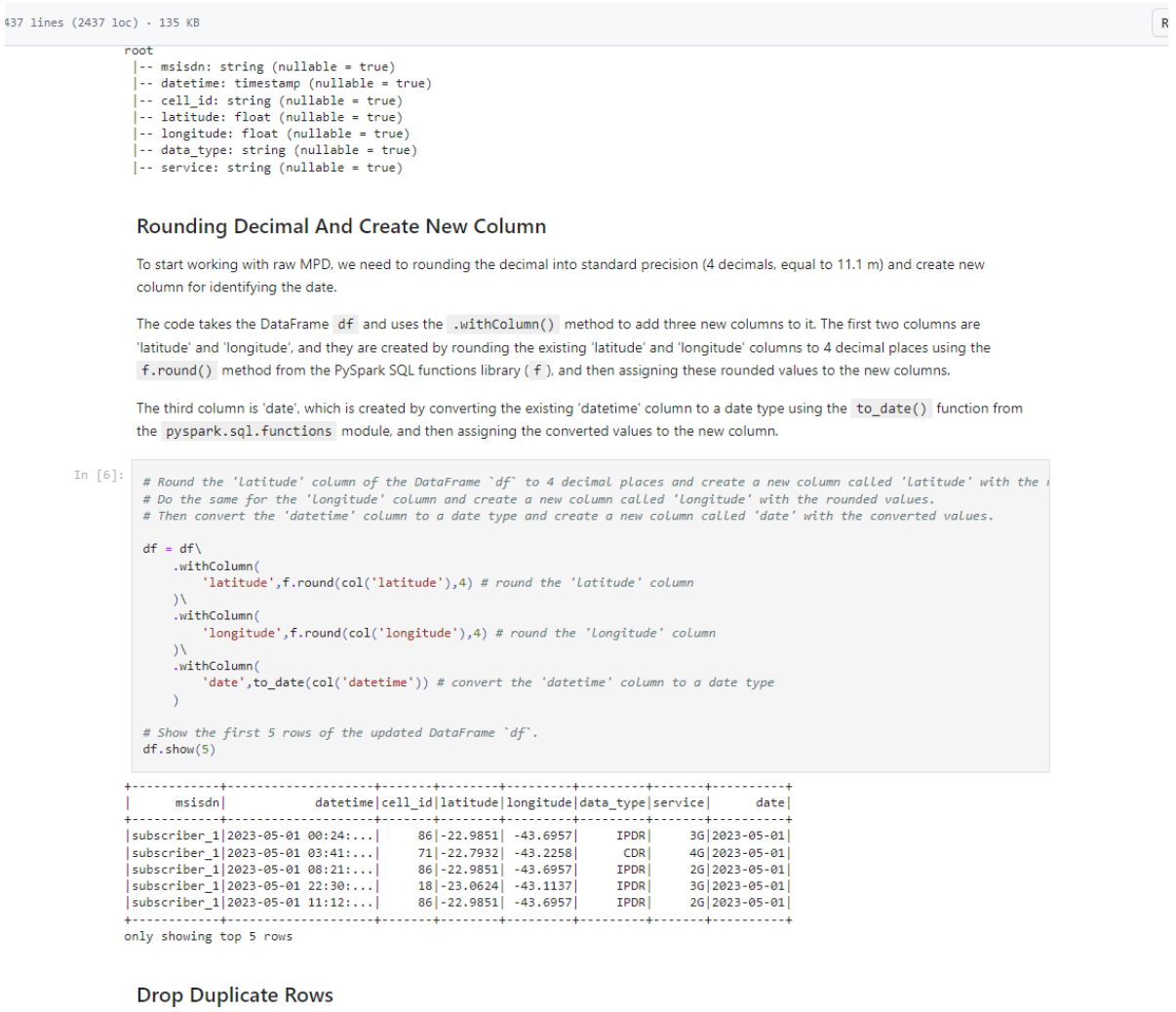

- ✓ Online
- ✓ Self-paced course
- ✓ Target population – anyone interested in mobile phone data

https://academy.itu.int/training-courses/full-catalogue/mobile-phone-data-awareness-course

- ✓ Content:
  - Opportunities and challenges
  - Data Access
  - Current applications

# Way forward

- Training materials for project managers
- Training materials for data scientists
- Generate synthetic data
- Python codes to help NSOs and agencies to calculate information society (SDG) indicators
- Deliver workshops and trainings (with the regional hubs and partners)
- Work with partners in integrating MPD as one of the data sources



437 lines (2437 loc) · 135 KB

```
root
 |-- msisdn: string (nullable = true)
 |-- datetime: timestamp (nullable = true)
 |-- cell_id: string (nullable = true)
 |-- latitude: float (nullable = true)
 |-- longitude: float (nullable = true)
 |-- data_type: string (nullable = true)
 |-- service: string (nullable = true)
```

### Rounding Decimal And Create New Column

To start working with raw MPD, we need to rounding the decimal into standard precision (4 decimals, equal to 11.1 m) and create new column for identifying the date.

The code takes the DataFrame `df` and uses the `.withColumn()` method to add three new columns to it. The first two columns are 'latitude' and 'longitude', and they are created by rounding the existing 'latitude' and 'longitude' columns to 4 decimal places using the `f.round()` method from the PySpark SQL functions library (`f`), and then assigning these rounded values to the new columns.

The third column is 'date', which is created by converting the existing 'datetime' column to a date type using the `to_date()` function from the `pyspark.sql.functions` module, and then assigning the converted values to the new column.

In [6]:

```
# Round the 'latitude' column of the DataFrame `df` to 4 decimal places and create a new column called 'latitude' with the r
# Do the same for the 'longitude' column and create a new column called 'longitude' with the rounded values.
# Then convert the 'datetime' column to a date type and create a new column called 'date' with the converted values.

df = df\
    .withColumn(
        'latitude',f.round(col('latitude'),4) # round the 'latitude' column
    )\
    .withColumn(
        'longitude',f.round(col('longitude'),4) # round the 'longitude' column
    )\
    .withColumn(
        'date',to_date(col('datetime')) # convert the 'datetime' column to a date type
    )

# Show the first 5 rows of the updated DataFrame `df`.
df.show(5)
```

```
+------------+--------------------+-------+--------+---------+---------+-------+----------+
|     msisdn|            datetime|cell_id|latitude|longitude|data_type|service|      date|
+------------+--------------------+-------+--------+---------+---------+-------+----------+
|subscriber_1|2023-05-01 00:24:...|     86|-22.9851| -43.6957|     IPDR|     3G|2023-05-01|
|subscriber_1|2023-05-01 03:41:...|     71|-22.7932| -43.2258|      CDR|     4G|2023-05-01|
|subscriber_1|2023-05-01 08:21:...|     86|-22.9851| -43.6957|     IPDR|     2G|2023-05-01|
|subscriber_1|2023-05-01 22:30:...|     18|-23.0624| -43.1137|     IPDR|     3G|2023-05-01|
|subscriber_1|2023-05-01 11:12:...|     86|-22.9851| -43.6957|     IPDR|     2G|2023-05-01|
+------------+--------------------+-------+--------+---------+---------+-------+----------+
only showing top 5 rows
```

### Drop Duplicate Rows

**Thank you!**
**indicators[at]itu.int**