8th International Conference on BIG DATA & Data Science for Official Statistics
BILBAO 2024

Informing Climate Change and Sustainable Development Policies with Integrated Data

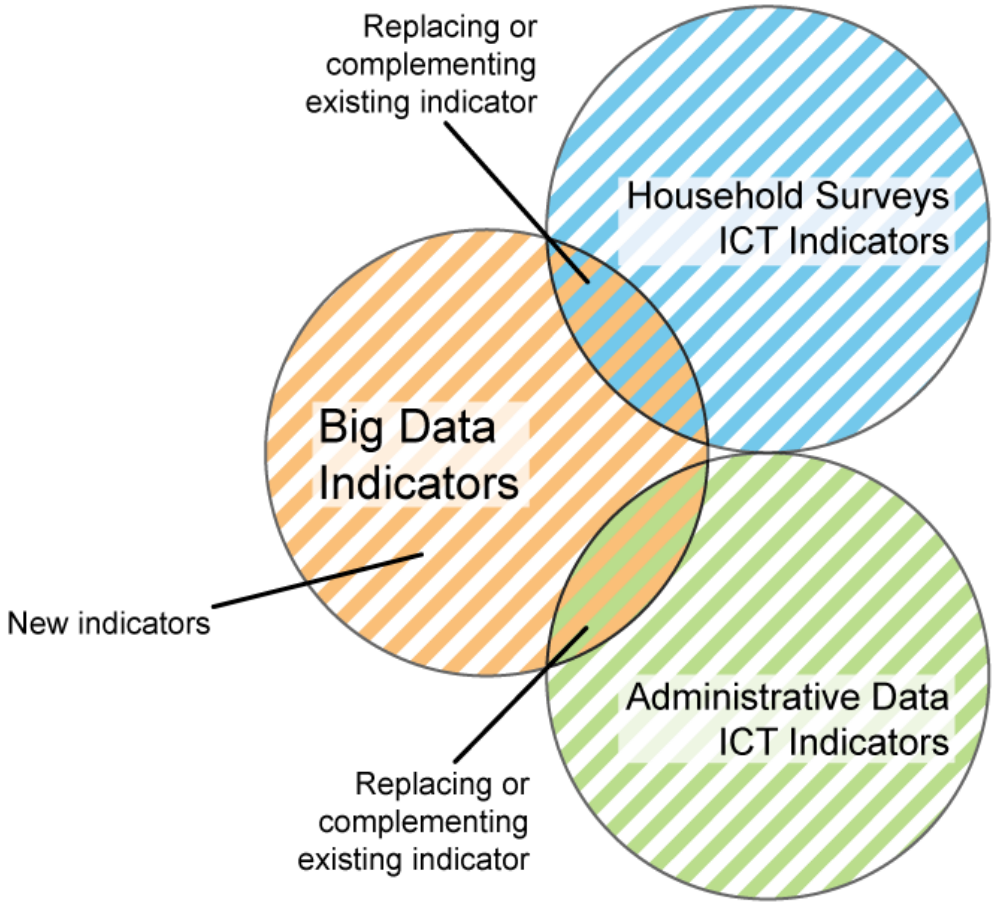
BILBAO. SPAIN | 10-14 JUNE 2024 | #UNBigData2024

Demo – What is mobile phone data, use cases and the typical process?

Esperanza Magpantay

Senior Statistician
International Telecommunication Union

ITU Statistics Work



1st pilot: 2016-2018

- ✓ 6 countries (Colombia, Georgia, Kenya, Philippines, Sweden, UAE)
- ✓ 16 ICT indicators (administrative data)

2nd pilot: 2020-2021

- ✓ Brazil, Indonesia
- ✓ 2 SDG ICT indicators
  - ✓ 9.c.1 – Percentage of population covered by mobile network: 2G, 3G and 4G and above (administrative data)
  - ✓ 17.8.1 – Percentage of population using the Internet (household survey data)

Introduction - ITU Work on Big Data for Measuring the Information Society

✓Started in 2014:	Measuring the Information Society Report
✓2014-2020:	Sessions in World Telecommunication/ICT Indicators Symposium; ITU Expert Groups on ICT Indicators (EGTI/EGH)
✓2016 -2018:	1st pilot project on big data for measuring the information society in 6 countries (Colombia, Georgia, Kenya, Philippines, Sweden, UAE) – 1st Guidelines material
✓2019 – present:	ITU active role in the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) ITU leading the Task Team on Mobile phone big data
✓2020 – 2021:	2nd pilot project (Brazil and Indonesia), 2nd Guidelines material

What is Mobile Phone Data (MPD)

Mobile Phone Data is defined as any type of mobile **network event** data that are stored by the **mobile network operator** (MNO) that includes a **subscriber identifier**, **time attribute** and **location**.

Call Detail Records (CDR)

- Passively generated when a subscriber:
  - Makes or receives a call
  - Sends or receives an SMS
  - Uses mobile data
- Routinely stored by MNOs for billing purposes

Passive Signalling Data

- Passively generated when a mobile device connects to the network
- More frequent than CDR data
- Storage depends on MNO

Active Signalling Data

- Mobile device locations determined by sending out an active 'ping'
- Used irregularly and at a significant cost to MNO
- Subscribers may opt-out from certain uses

Statistics areas where MPD can be used

1) Tourism statistics

2) Migration statistics

3) Census and dynamic population

4) Displacement in disaster

5) Information society indicators

6) Transport and commuting statistics

Assumptions:

- Necessary data (MPD and reference data) are available
- Hardware and software tools are available
- Necessary skills are available

Data requirements:

1. Data from mobile network operators
2. Reference data
   - Local Administrative Units (LAU)
   - World Population
   - Cell location data
   - Digital elevation model
   - Household surveys and microdata

A Team of Experts: Diverse skills

The team to work on MPD should normally be composed of these and other staff to make sure to cover all the necessary skills

Sys-op for managing the server, software and file system (where MPD, cells and reference data, etc. are stored)

Data engineers who prepare reference data, raw data files and set up configuration

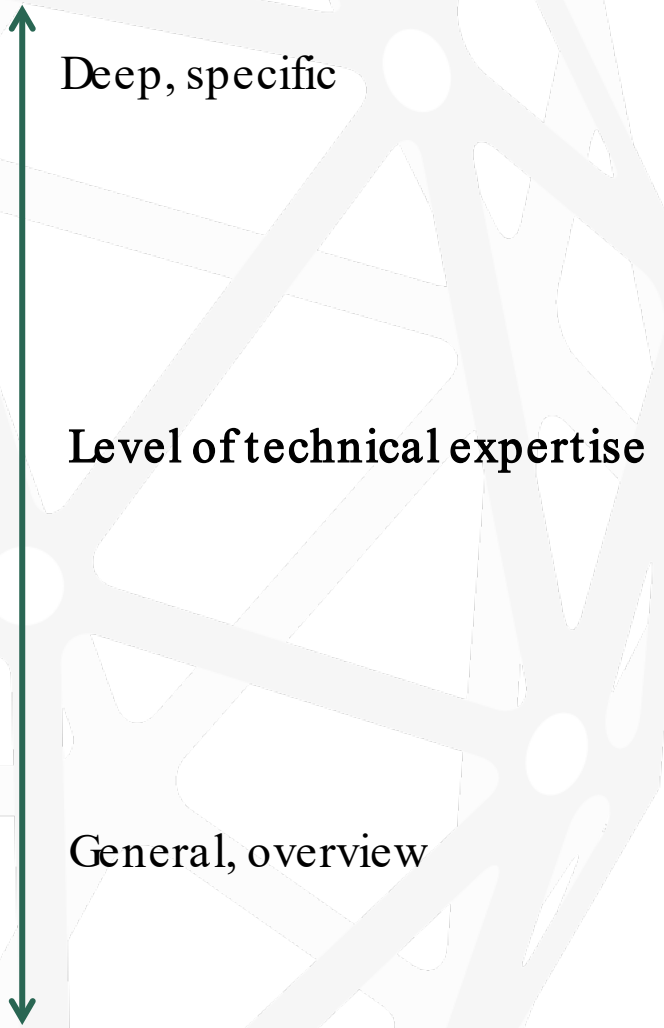
GIS specialist who conducts QA on reference data (administrative units, roads, etc.)

MPD QA specialist who prepares the MPD files received from MNOs and conducts data QA

Domain methodology experts who consult on configuration files and methodology adjustment

Experienced Project Manager to manage all those activities

Necessary technical skills from one example MPD project that mainly used PostgreSQL:
Linux OS and file system, bash and awk, Python3, SQL, PostgreSQL10+, PostGIS2.5+, Apache2, PHP, RESTful API, understanding data security principles and secure data transfer methods



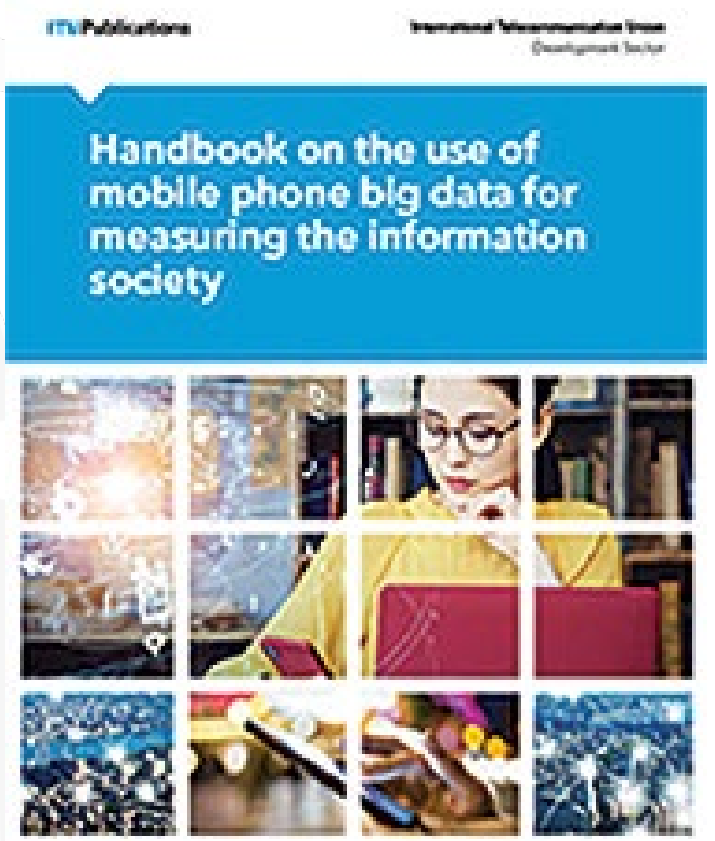
Process of analyzing MPD

1. Data quality checks of raw data
2. Processing of raw data (centralized, distributed, PPP)
   Ensuring privacy and data protection
3. Establishing the home location
4. Calculation of indicators
5. Visualizations
6. Quality assurance of calculated data

Methodological Guide on Big Data for measuring the SDG Information society indicators

1. Introduction
2. Background
3. Access and preparations
4. Data sources (description of mobile operator data, quality assurance of raw data)
5. Reference data (local admin units, world population, cell data, digital elevation, household survey data)
6. Data processing (models, data protection guidelines)
7. Calculating the indicators (rationale, definition, indicators calculation, quality assurance)
8. Quality assurance
9. Conclusions

- with experiences and examples from country pilots



Next: Processing of raw MPD

Thank you!

More information: magpantay[at]itu.int

#UNBigData2024