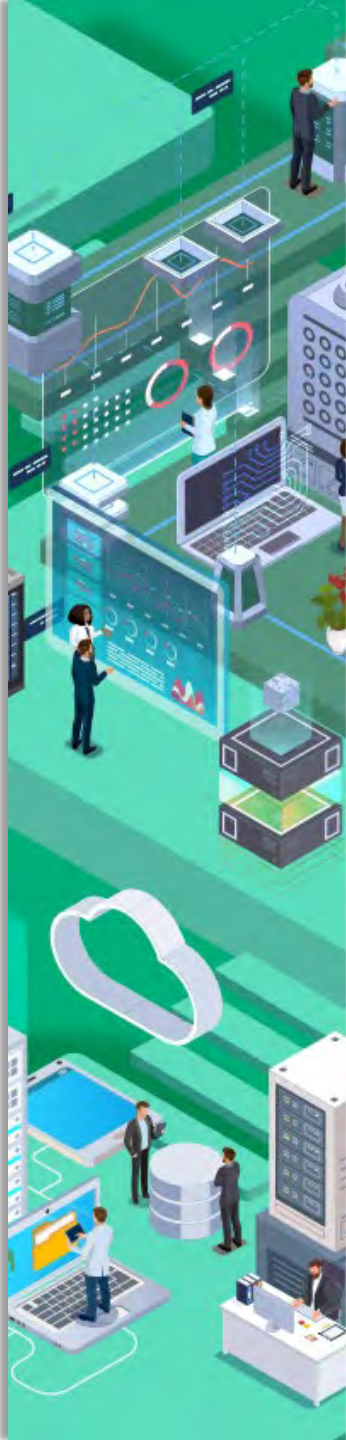


# ITU big data for measuring the information society

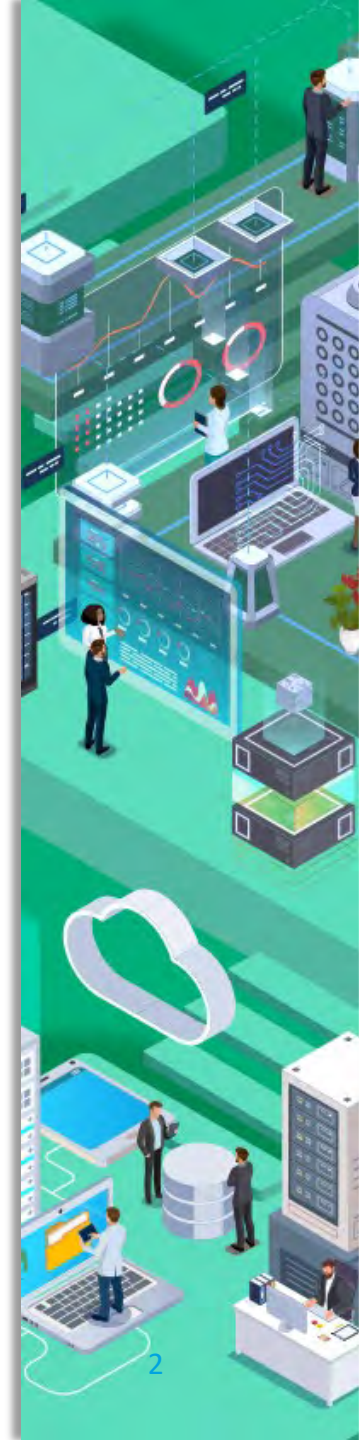
1. Using Mobile Phone Data (MPD)

2. Using open-source data



# Innovation in Data Collection for SDGs

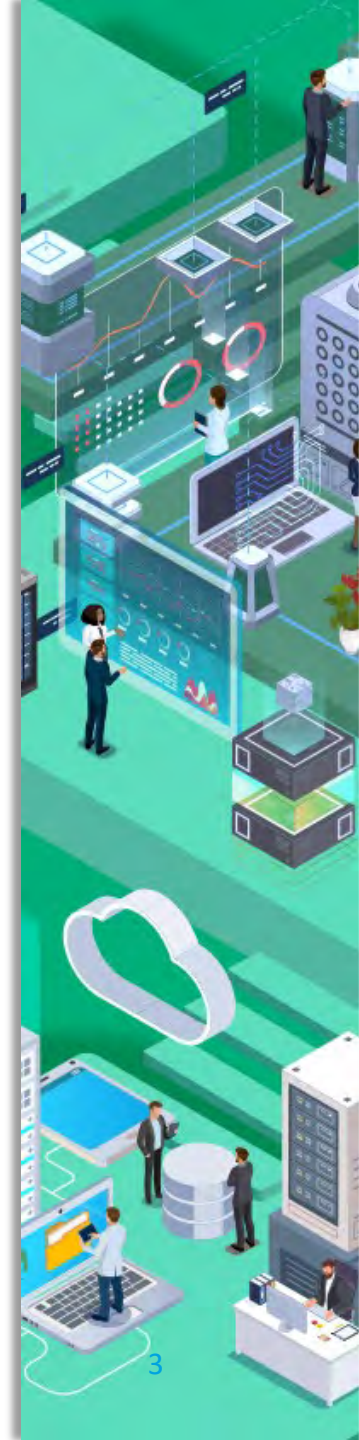
- National statistical systems are under increasing **pressure** to **produce high-quality** data to monitor **SDGs**.
- Recent challenges in statistics production:
  - the need for faster and more timely statistics
  - the need for disaggregated information
  - the rise of new data sources



# Big Data Sources

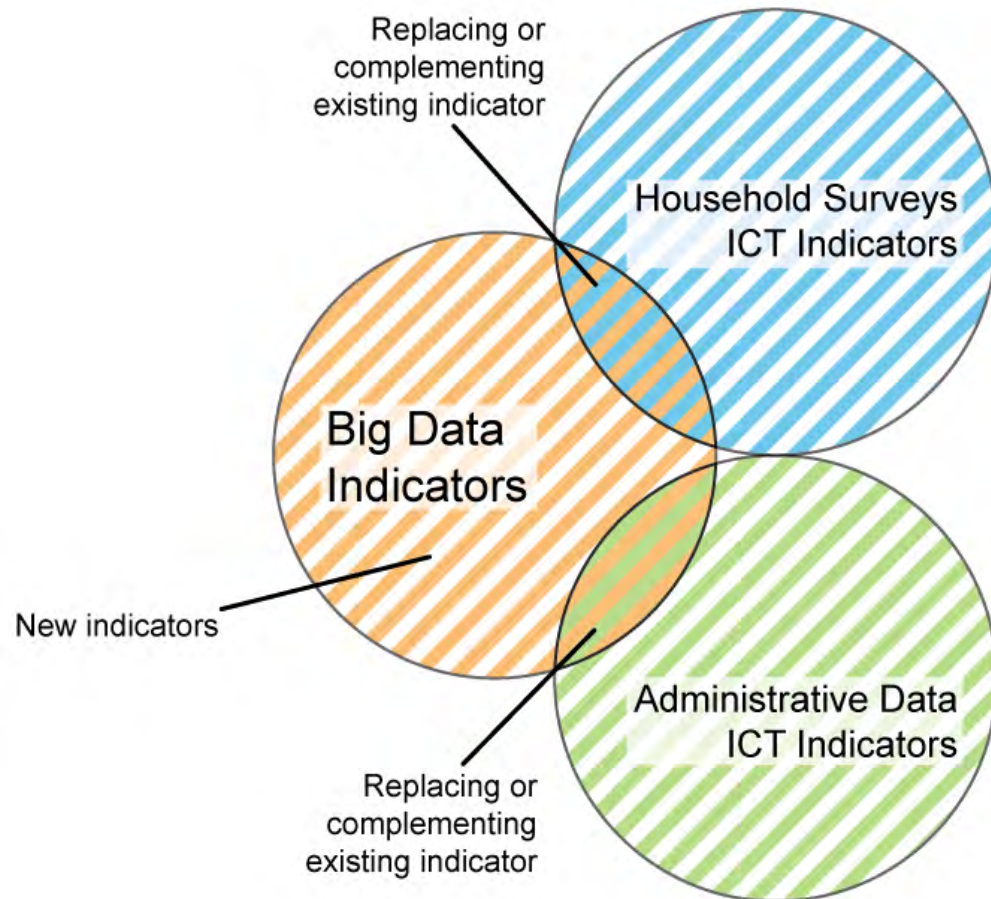
- Telecommunication Networks (Mobile phone data)
- Social Media
- Instant messaging
- Web crawling and scraping
- Sensors
- Remote sensing
- IoT (machine to machine)
- Human-generated bio-data
- Mobile payment platforms
- .....

Big data nexus	Amount of information retrieved	Relevance	Simplicity	Coverage	Timeliness	Continuity	Practical feasibility and resources required
Telecommunication network	••••	••••	••••	••••	••••	•••	•
Mobile devices	••••	••••	•	•••	•••	••	•••
Social networks	••••	•••	•••	•••	•••	•	••
Instant messaging services	•	••••	••	•••	•••	•	••
Telecommunications equipment	••••	••••	•	•••	•••	••	••
Auto-update services	••	••	••	••••	••	••	•••
CDNs	••••	••••	••	•••	••••	••	•••
IaD network-centric	••	••	•••	•••	•••	••	••••
IaD site-centric	•	•	•	•••	•••	•••	•••
IaD crawling & scraping	••••	•••	•••	••	•••	•••	•••
Spamhaus	••	••	•••	•••	••	••	•••
Virus scanners	•	•	••	••	••	•	••
Mobile payment platforms	•••	••	•	••••	•••	•	•





# ITU Statistics Work



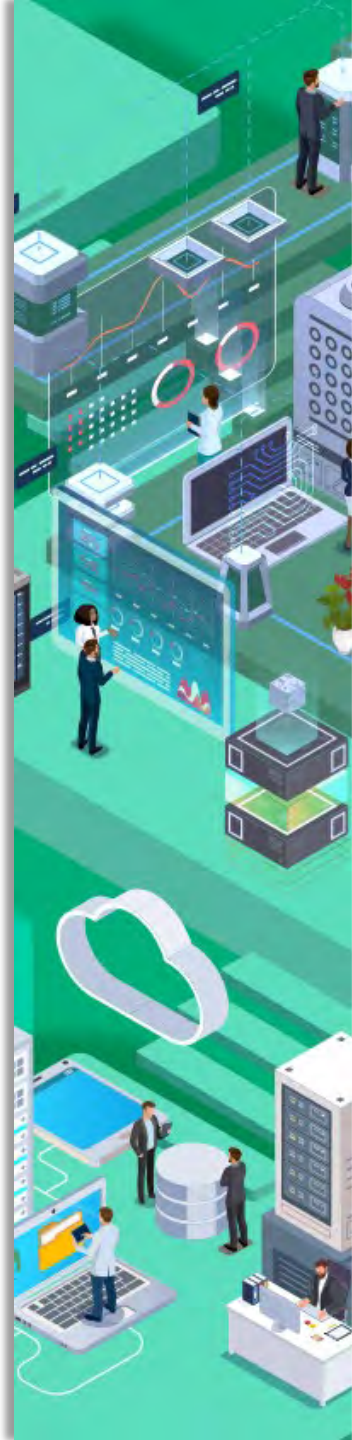
## 1st pilot: 2016-2018

- ✓ 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)

<https://www.itu.int/en/ITU-D/Statistics/Pages/bigdata/default.aspx>



# 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)

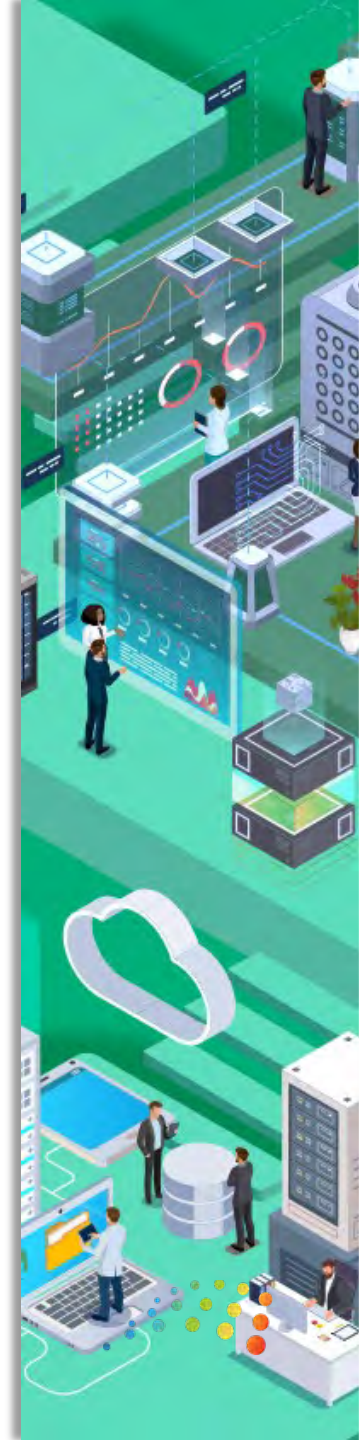
- Passively generated when a subscriber:
  - Makes or receives a call
  - Sends or receives an SMS
  - Uses mobile data
- Routinely stored by MNOs for billing purposes

## 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



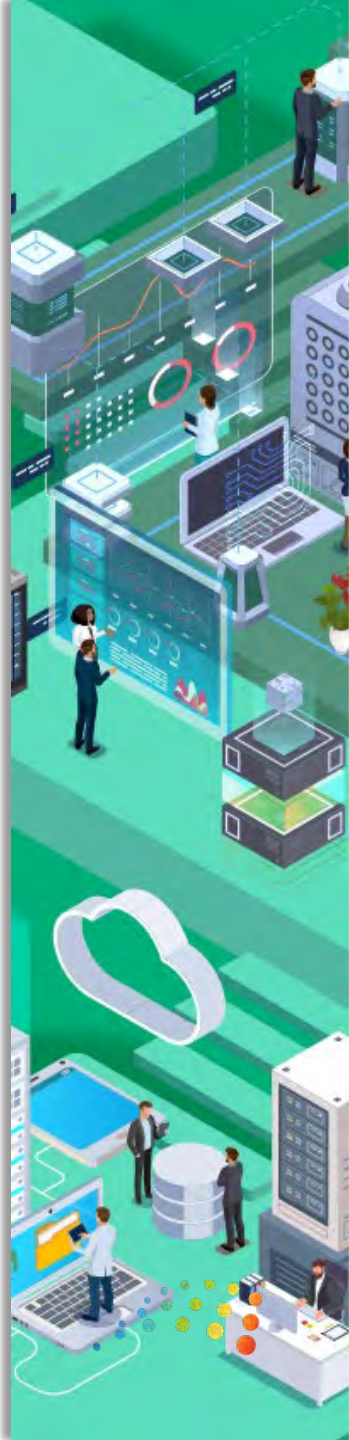
# What are the strengths and limitations of MPD?

## Strengths

- High penetration globally
- High quality data
- Near-real-time
- High granularity
- Already generated by organisations

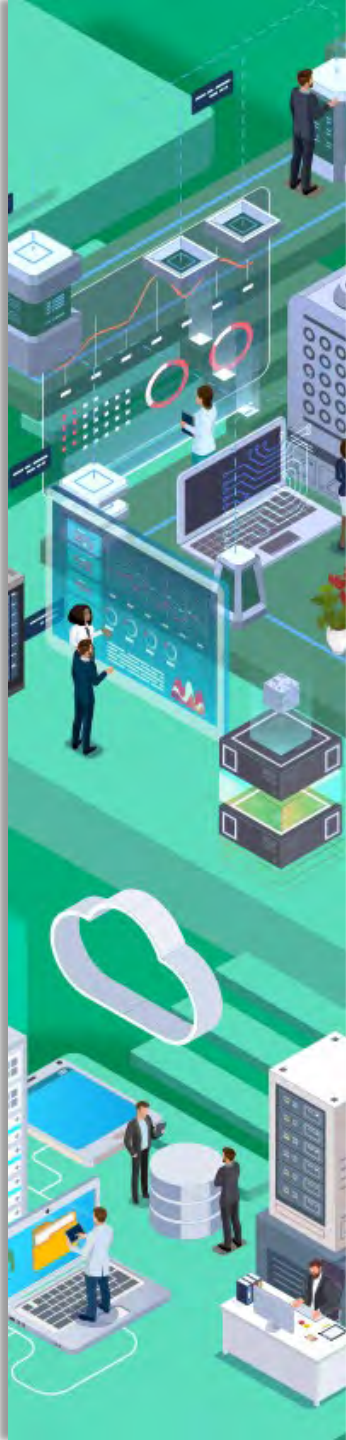
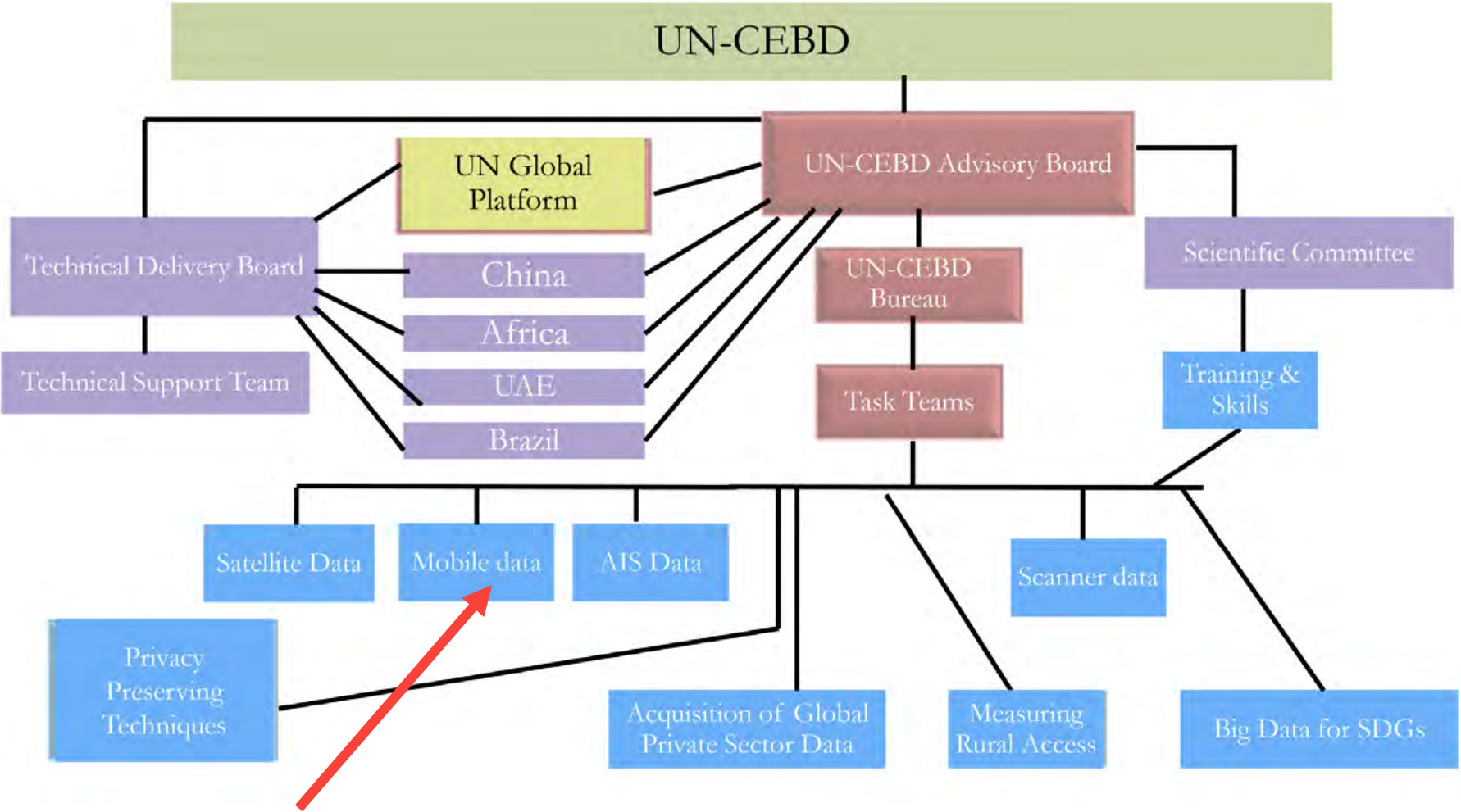
## Limitations

- Precision limited by cell tower density
- Representativeness of mobile phone subscriptions
- One subscription may not represent one individual, and vice versa
- Access requires agreement with MNOs
- Large datasets requiring suitable data infrastructure





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



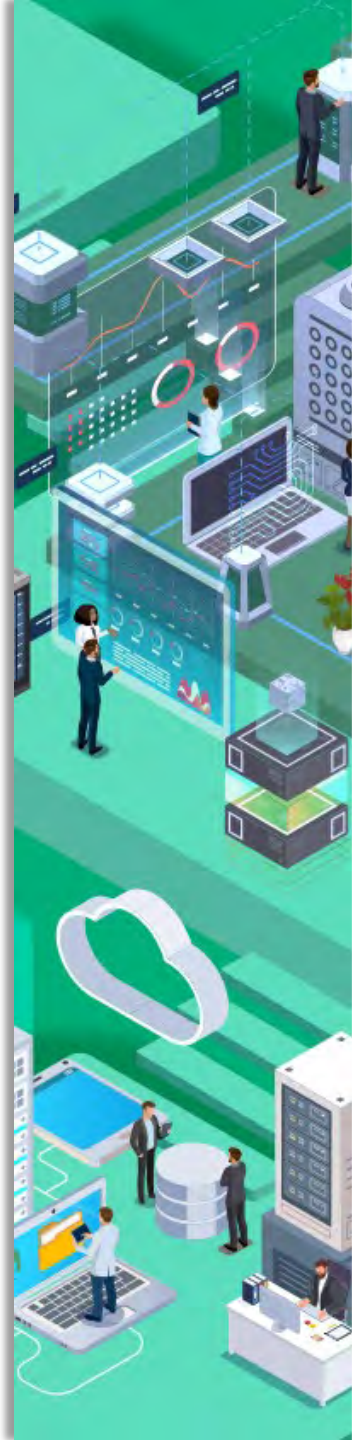
# UN-CEBD Task Team on Mobile Phone Data (MPD)

- Established since the inception of the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)
- Explore the use of mobile phone big data for the different areas of statistics
- Composed of around 50 individual members/ 30 entities - international and regional agencies, countries, academia, private agencies/companies
- Meets (virtually) once a month to discuss issues related to the deliverables, events, and other activities

## Members:

### ITU (Chair)

- Brazil
- Colombia
- Georgia
- India
- Indonesia
- Italy
- Japan
- Korea
- Malaysia
- Mexico
- Netherlands
- Oman
- Philippines
- Romania
- Saudi Arabia
- United Arab Emirates
- EU JRC
- Eurostat
- IMF
- IOM
- UNFPA
- UNGP Jakarta
- UNSD
- World Bank
- OECD-ITF
- UN-ECE
- Flowminder
- GSMA
- Positium
- Telenor





# Five key principles for maintaining public trust

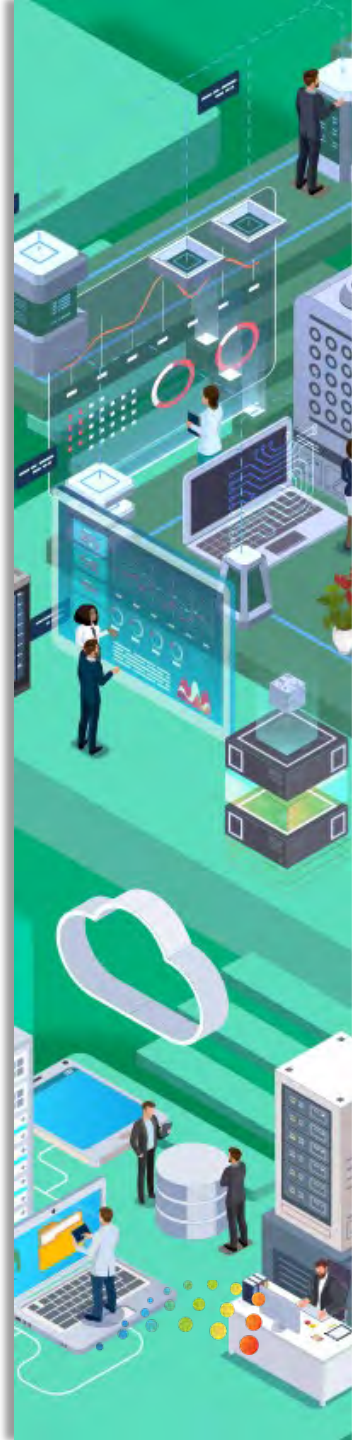
- when using of mobile phone data:

1. Necessity and proportionality
2. Professional independence
3. Privacy-conscientiousness
4. Commitment to quality
5. International comparability

In general, processing MPD for statistical purposes shall be subject to:

- appropriate safeguards,
- privacy best practices, and
- relevant laws, including GDPR, to protect the rights and freedoms of the data subject.

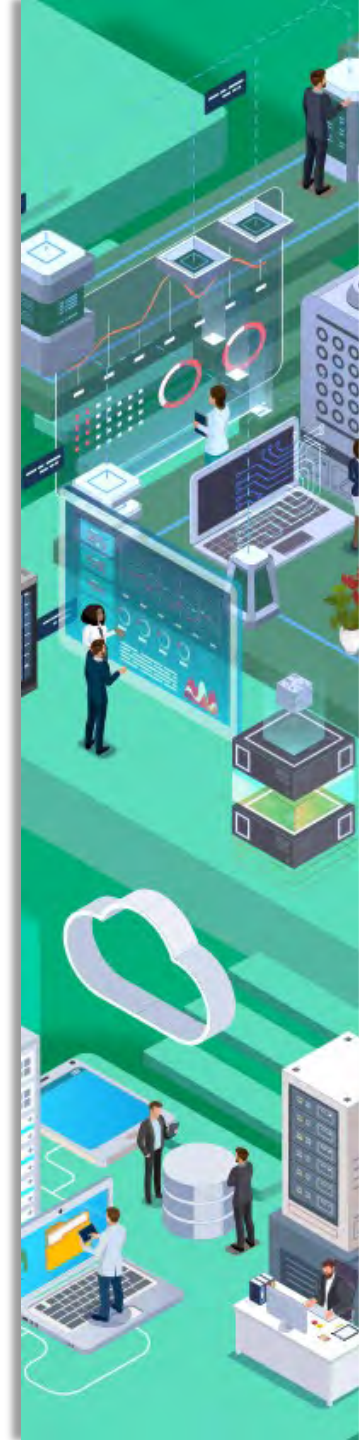
Those safeguards shall ensure that technical and organizational measures are in place to comply with the principle of data protection. Those measures may include, but are not limited to, pseudonymization. Pseudonymized data remains personal data, and relevant data protections laws continue to apply.



# Task Team on Mobile phone data - sub-groups

Areas of statistics where mobile phone data can be used:

- 1) Tourism statistics (lead: BPS Indonesia)
- 2) Migration statistics (Lead: GeoStat, Georgia)
- 3) Census and dynamic population (lead: Positium)
- 4) Transport and commuting statistics (lead: World Bank)
- 5) Information society indicators (**lead: ITU**)
- 6) Displacement in disaster context (lead: University of Tokyo)



# Requirements for MPD project

Objective  
Statement

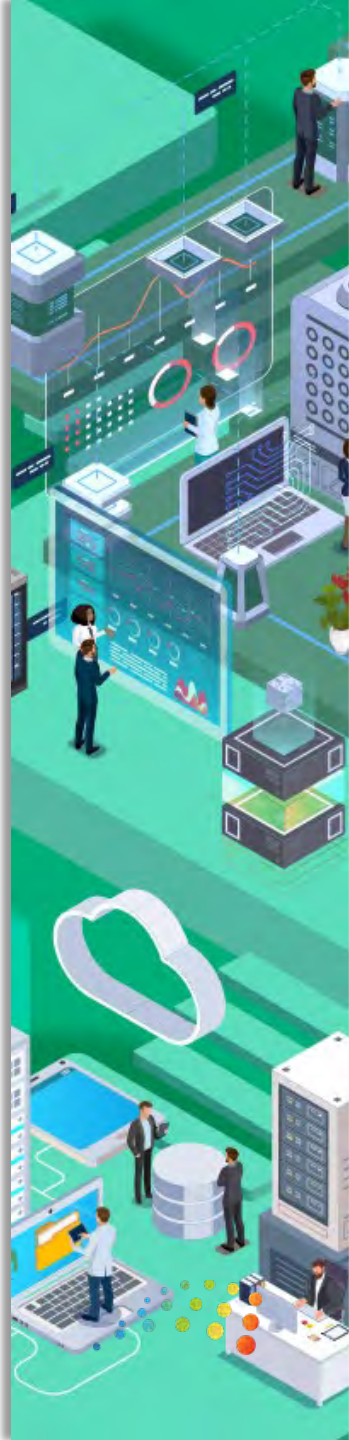
Access to  
MPD

Project  
Sponsor

Data  
Scientists

Software  
Tools

Hardware  
(Servers)

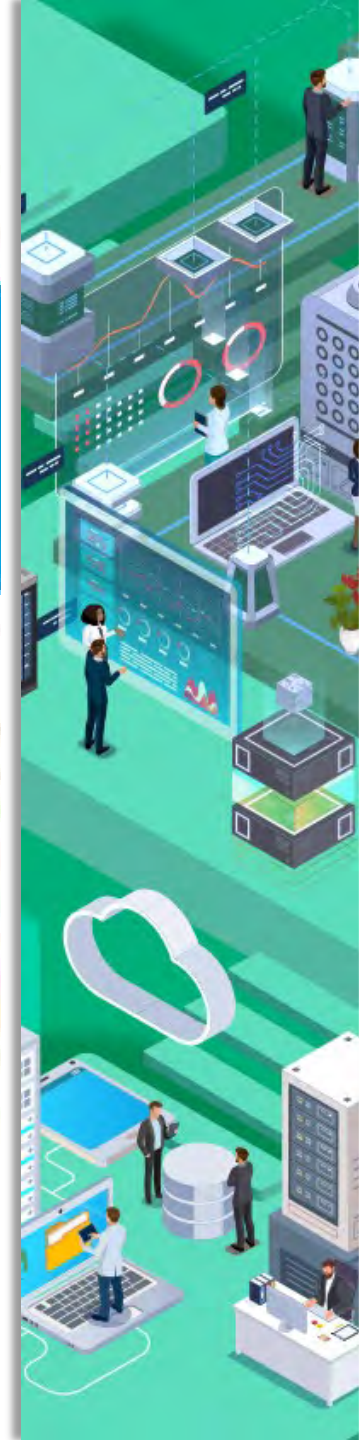




# Guidelines on MPD for measuring the Information society

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

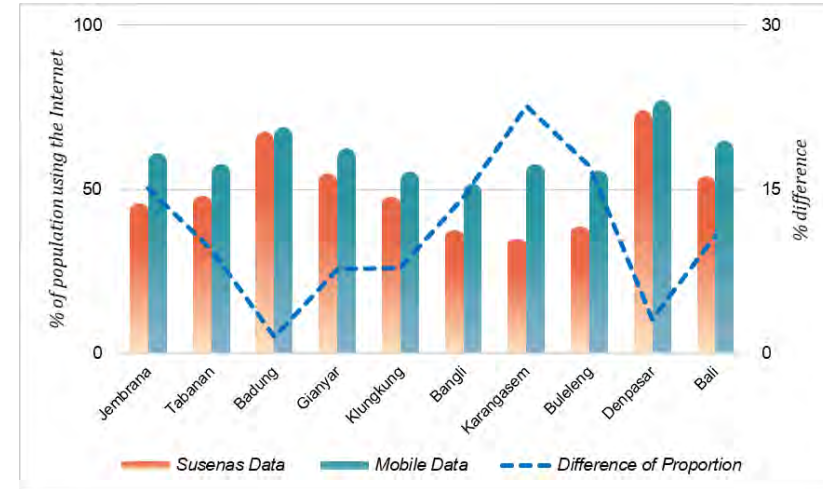
- with experiences and examples from country pilots



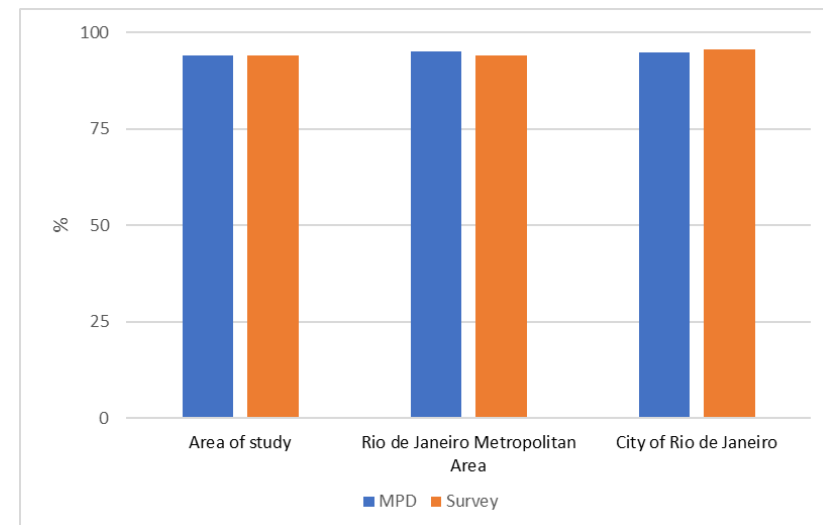
# Measuring Internet use in Brazil and Indonesia, 2021

- MPD used for **ICT statistics** to assess progress against SDG indicators such as **proportion of individuals using the Internet (17.8.1)**
- Used in both Indonesia and Brazil
- Comparison to existing household survey methodology shows **consistency** between estimates produced by MPD and traditional data sources (household survey)

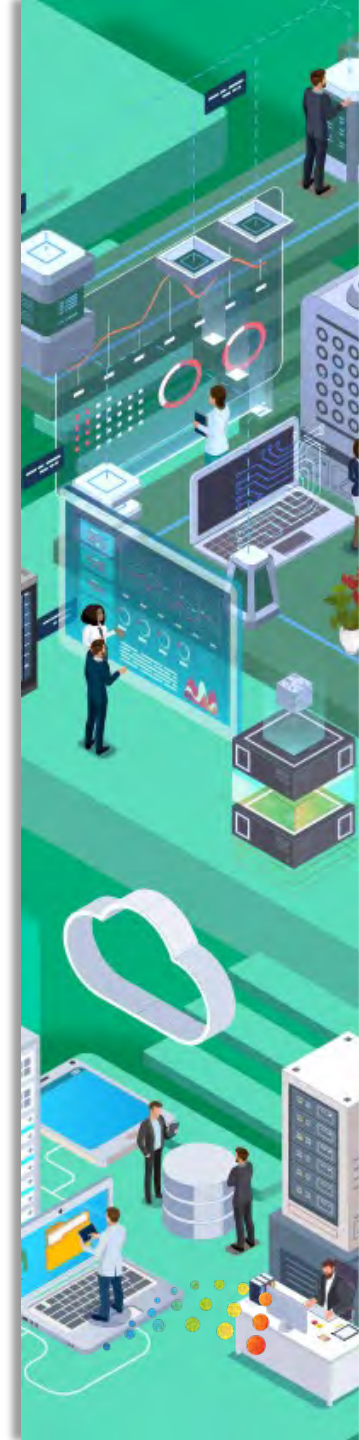
Indonesia



Brazil

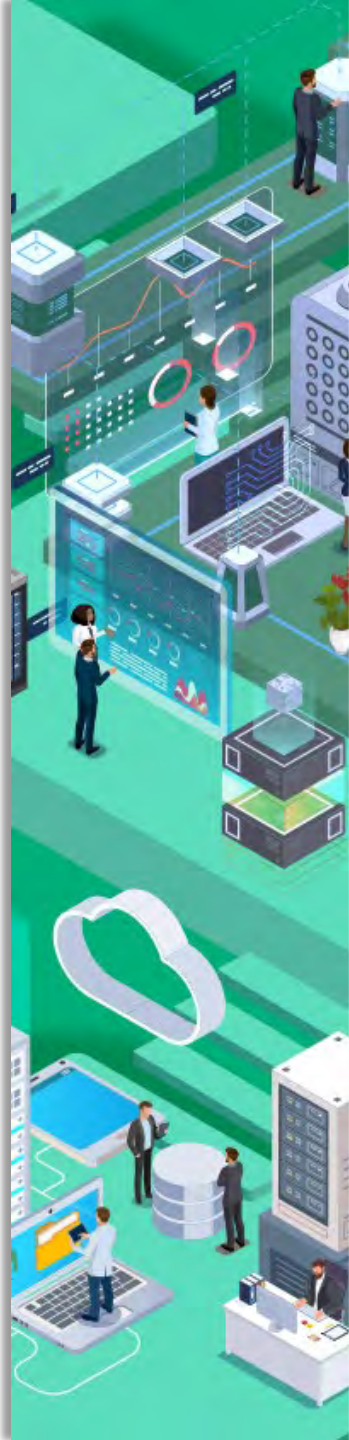


Source: ITU.



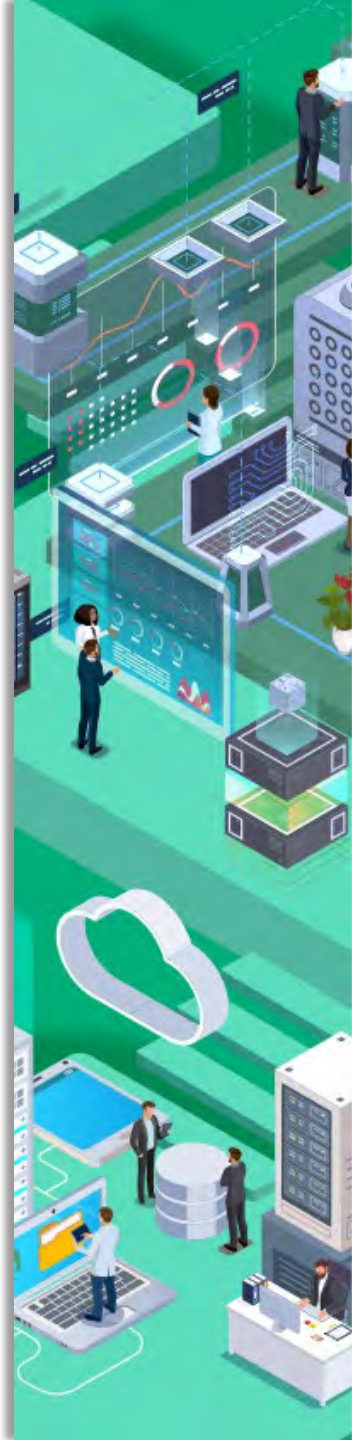
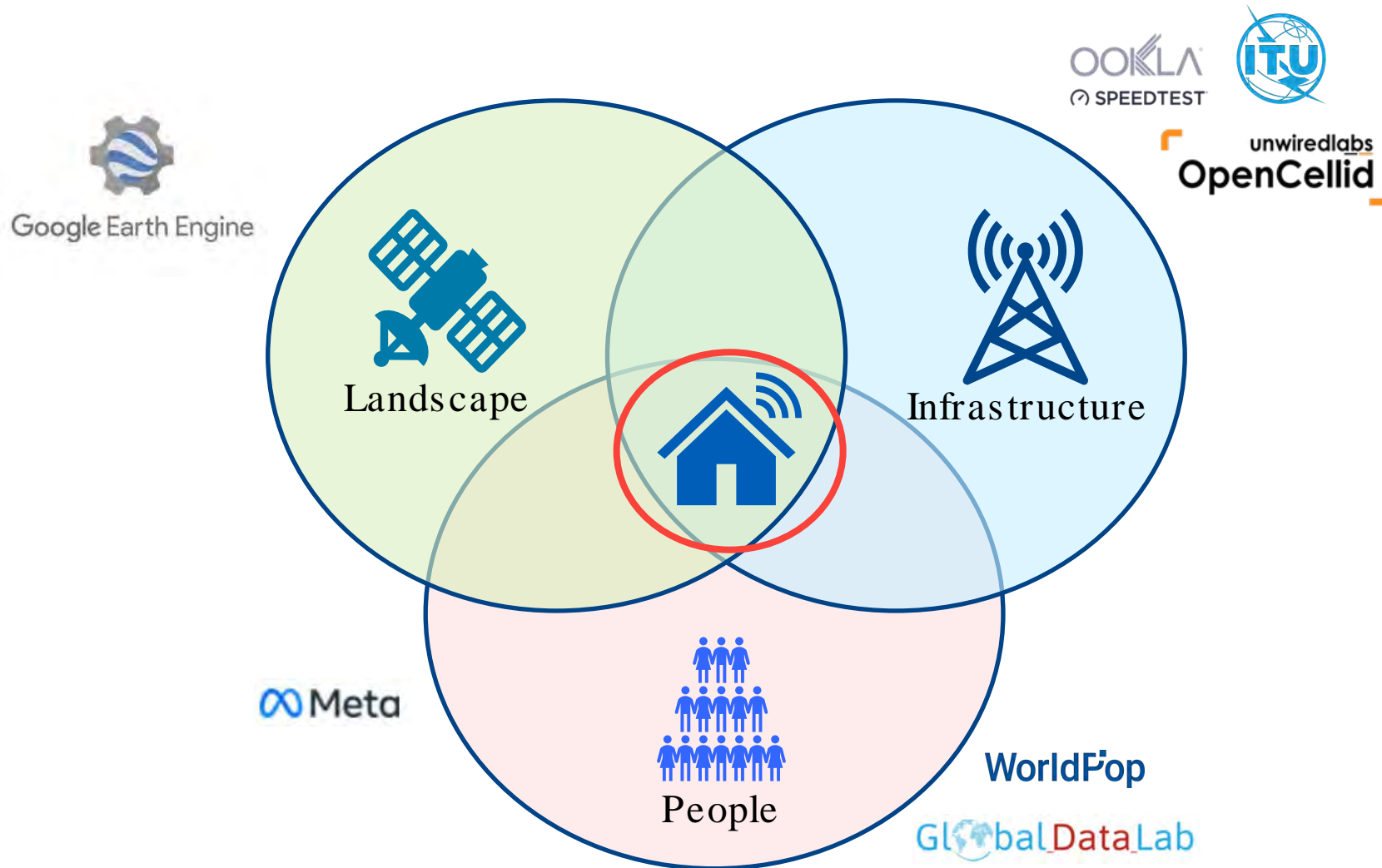
# Using mobile phone data - useful links

- ITU Big Data for Measuring the Information Society - <https://www.itu.int/en/ITU-D/Statistics/Pages/bigdata/default.aspx>
- UN-CEBD Task Team on MPD - <https://unstats.un.org/bigdata/task-teams/mobile-phone/index.cshtml>
- Paper "Guiding principles to maintain public trust in the use of mobile operator data for policy purposes" to be published in the Data and Policy Journal <https://www.cambridge.org/core/journals/data-and-policy>





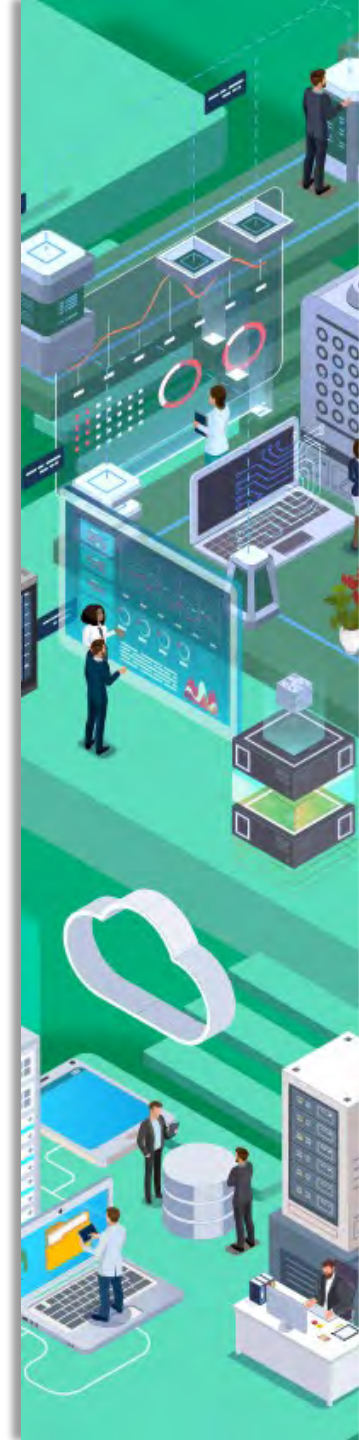
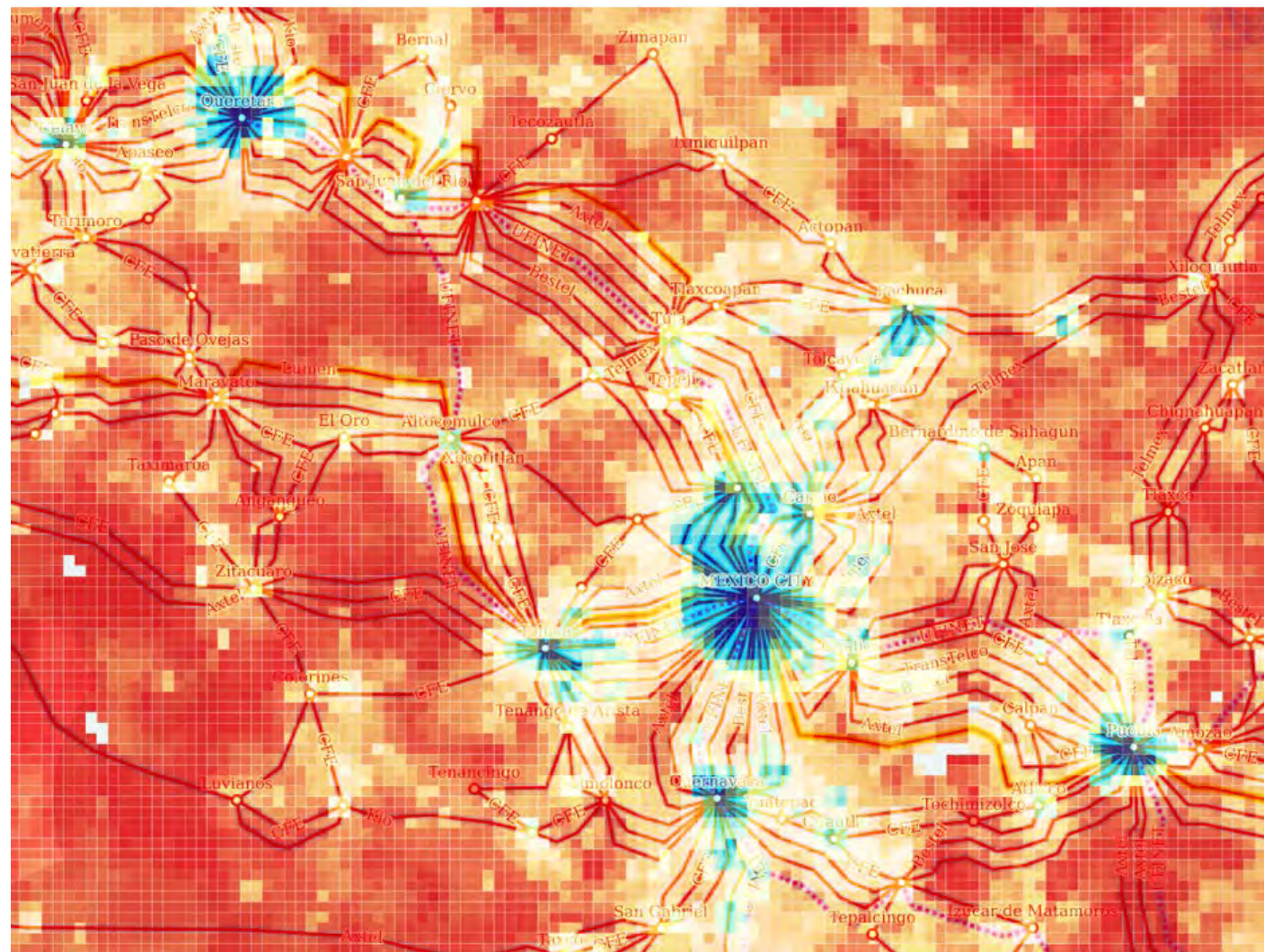
# Can open-source data bring us deeper insights on Internet use?





# “Pixelate” the country and map data

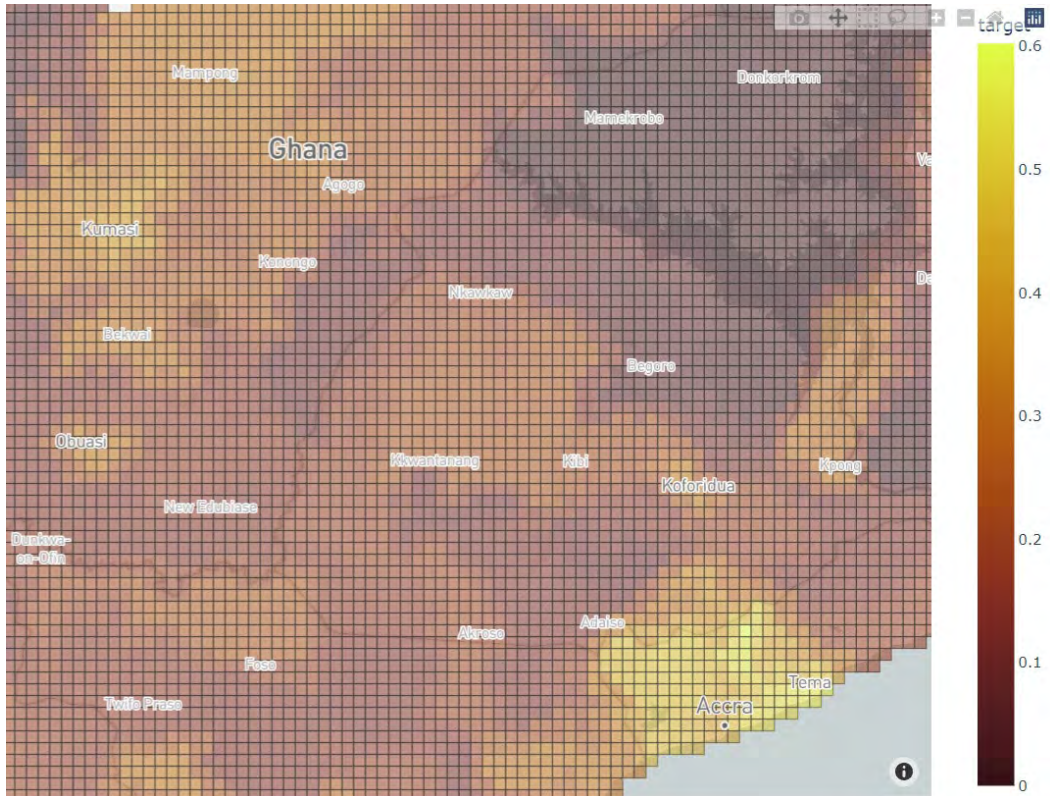
- Allow aggregation at various geographical levels
- Allow identification of “pockets” of populous, unconnected areas
- Can be mapped with specific locations, e.g. schools.



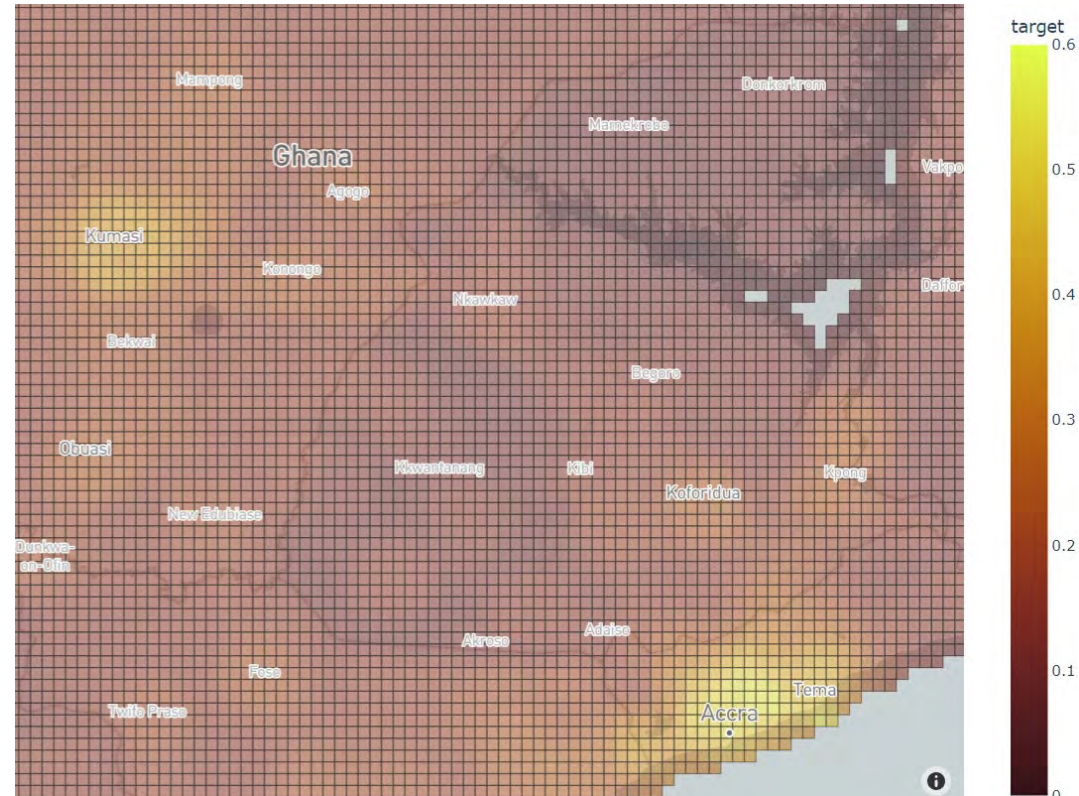


# Model development and validation

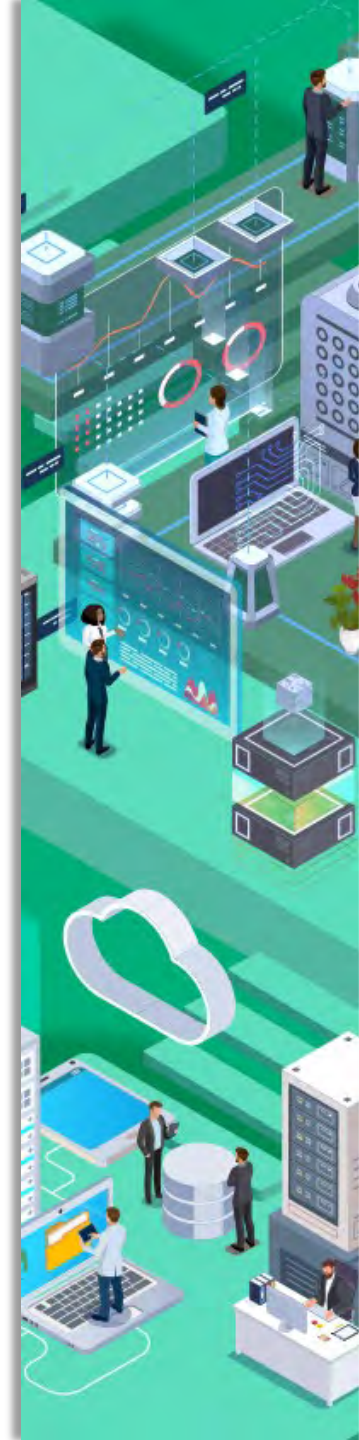
Actual Internet use



Predicted Internet use



Aggregated test data within **1 %** of actual



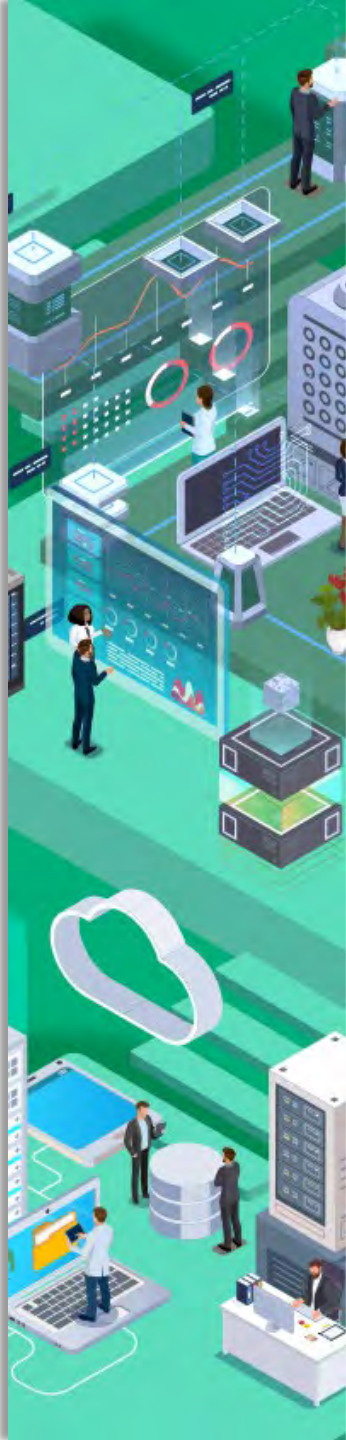


# Next steps

Refine  
model

More  
household  
data

Multi-  
country  
models





**Thank you!**

