Alternative data sources (Big Data) and innovative statistical methods

Viviana Umpierrez
Statistician
ICT Data and Analytics Division
International Telecommunication Union (ITU)

Mobile phone data

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

Overview – ITU mobile phone big data pilot projects

Objectives:
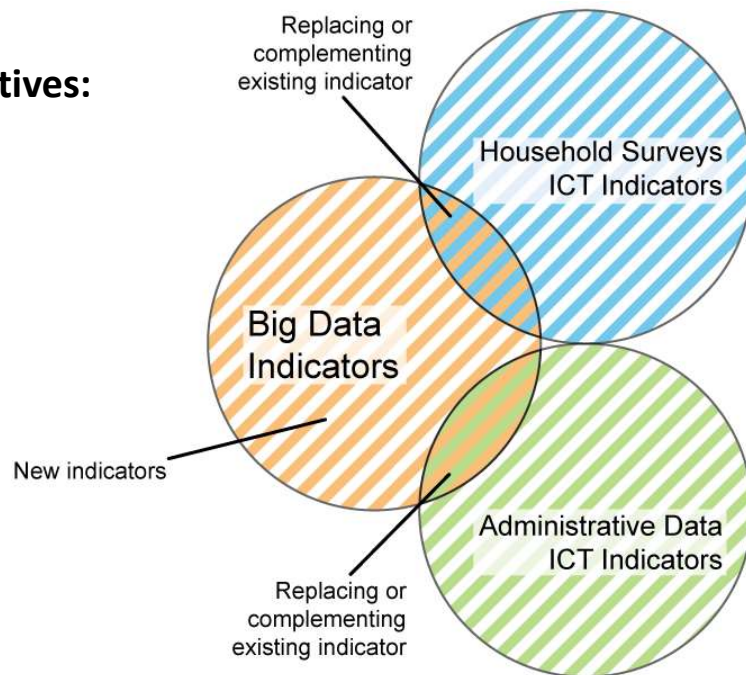
1st pilot: 2016-2018
- ✓ 6 countries (Colombia, Georgia, Kenya, Philippines, Sweden, UAE)
- ✓ 16 ICT indicators (administrative data)

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)

Challenges:
- ➢ Lack of skills to use big data
- ➢ Data access
- ➢ Lack of infrastructure

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

1st pilot

Stakeholders

- Telecommunication Regulator / ICT ministry
- National Statistics Office
- Telecommunication Service Providers (MNOs, ISPs)
- Data Protection Authority

Pilot Countries

- Colombia
- Georgia
- Kenya
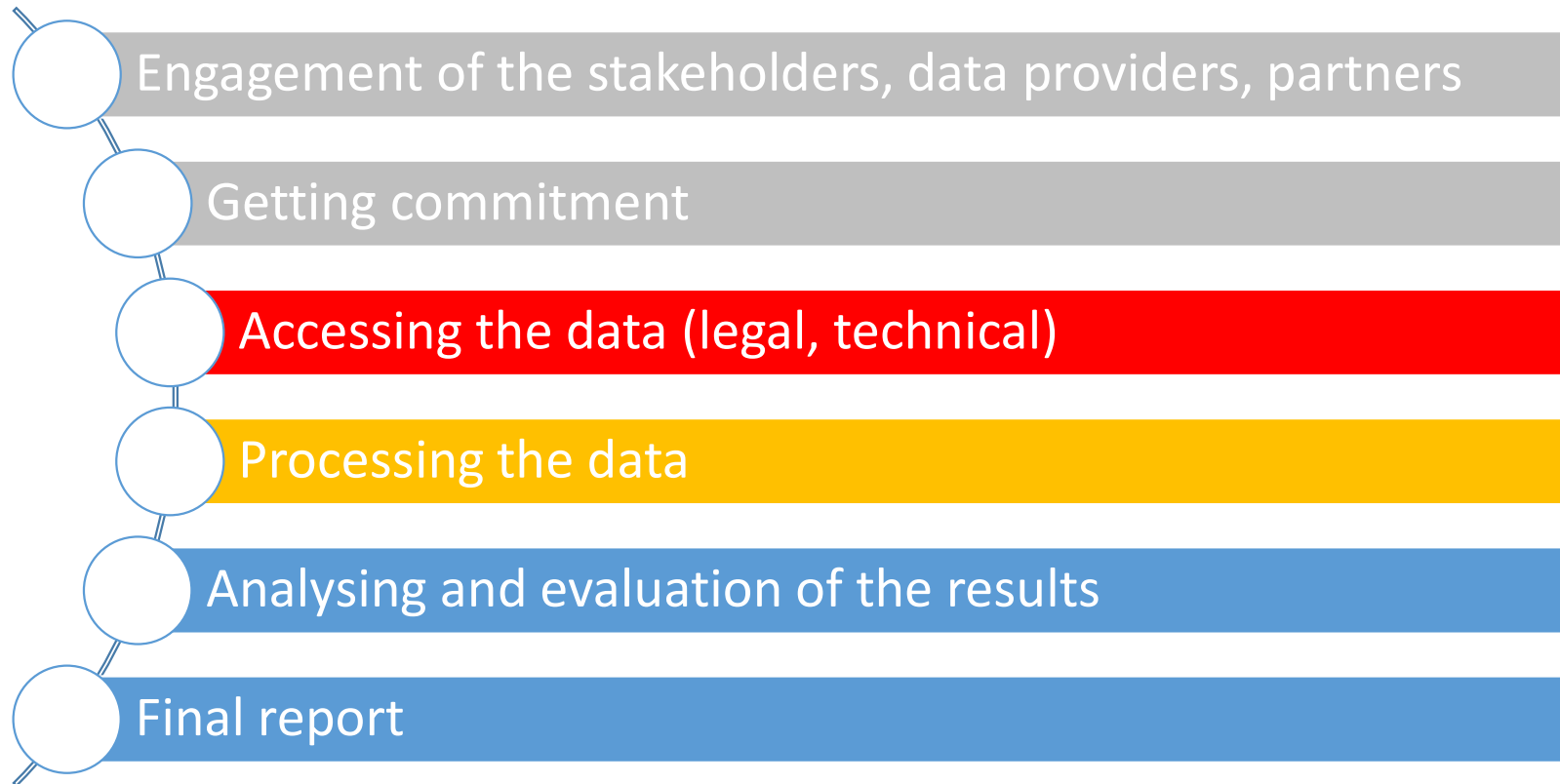- Philippines
- Sweden
- United Arab Emirates

- Two data scientists and project coordinator to assist pilot countries

Roles of different stakeholders

1. *Telecommunication Regulator / ICT ministry*
   - ✓ Request mobile phone data as a requisite for monitoring licensing condition
   - ✓ Invested in equipment and expertise to store and process these records
   - ✓ Frequently interact with operators in the course of their regulatory work
   - ✓ Positioned to negotiate and mediate access to mobile phone data
2. *National Statistics Office*
   - ✓ Statistical Act
   - ✓ Mandate to produce official statistics and collect the data
   - ✓ Skills to analyse statistical information
3. *Mobile Phone Operators and service providers*
   - ✓ Custodian of mobile phone data
   - ✓ Invested in equipment and expertise to store and process these records
   - ✓ Required to submit records to a regulatory agency as a condition of their licence or franchise
   - ✓ Have staff that have big data skills to analyse MPD
4. *Data Protection Authority*
   - ✓ Provides guidance and oversight for lawful data processing
   - ✓ Ensures safeguards are in place to ensure privacy (pseudonymisation or anonymisation)

Stages of Implementation - June 2016 – November 2017

- Engagement of the stakeholders, data providers, partners
- Getting commitment
- Accessing the data (legal, technical)
- Processing the data
- Analysing and evaluation of the results
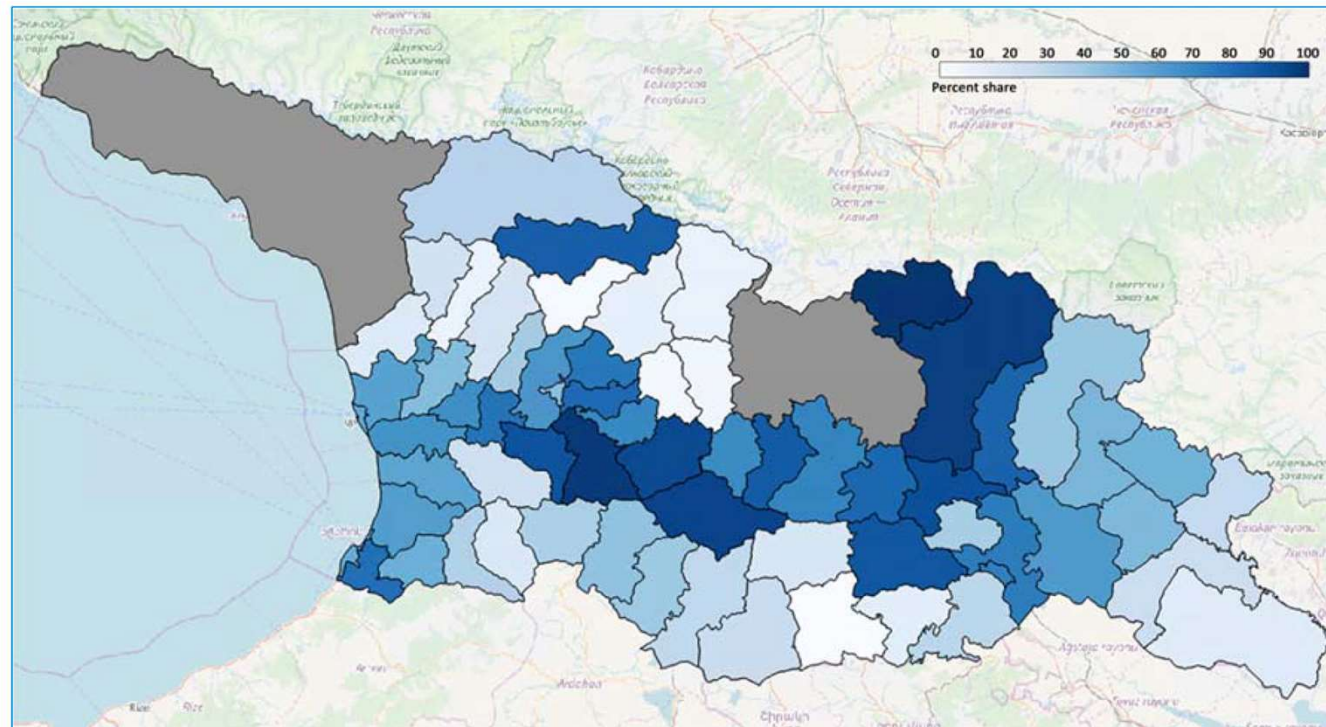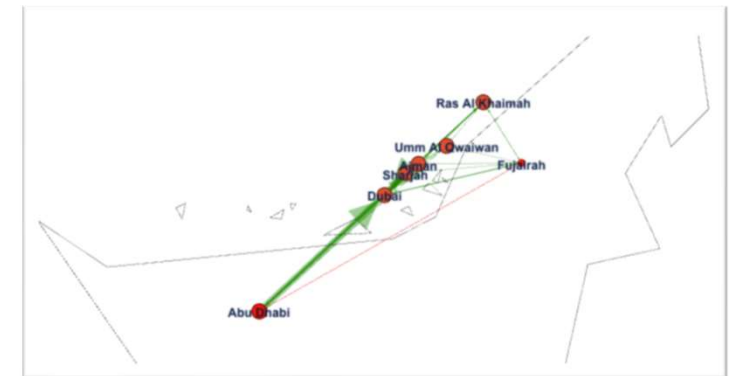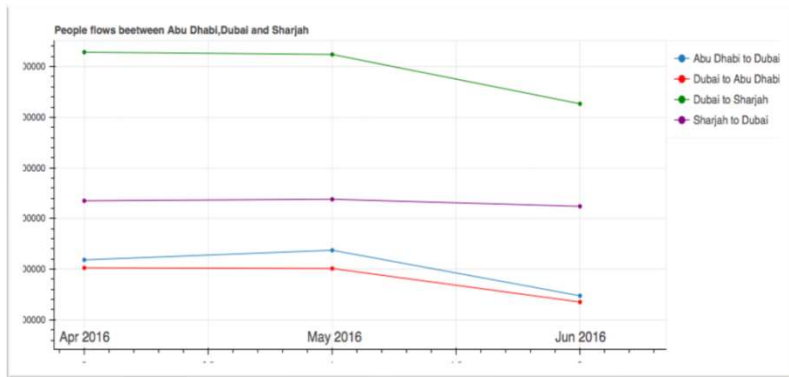- Final report

Big data indicators

- BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

- BD02: Percentage of the Population Covered by a Mobile-Cellular Network, by Technology

- BD03: Usage of Mobile-Cellular Networks for non-IP Related Activities, by Technology

- BD04: Usage of Mobile-Cellular Networks for Internet Access, by Technology

- BD05: Number of Subscriptions with Access to Technology

- BD06: Active Mobile Voice and Broadband Subscriptions, by Contract Type

- BD07: Average Number of Active Mobile Subscriptions per Day, by Contract Type

- BD08: Active Mobile Devices

- BD09: IMEI Conversion Rate

- BD10: Fixed Domestic Broadband Traffic, by Speed, Contract Type

- BD11: Mobile Domestic Broadband Traffic, by Contract Type, Technology

- BD12: Mobile International Broadband Traffic, by Contract Type

- BD13: Inbound Roaming Subscriptions per Foreign Tourist

- BD14: Fixed Broadband Subscriptions, by Technology

- BD15: Fixed Broadband Subscriptions, by Speed

- BD16+: Proposed New Indicators from Pilot Countries
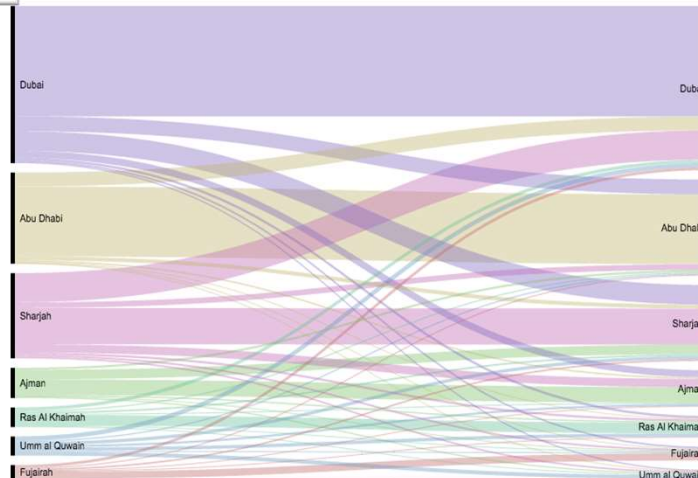
Georgia - results

Figure 39: Number of active smartphone devices in LAU2 districts

United Arab Emirates – results
BD16: Human Mobility

People flows beetween Abu Dhabi, Dubai and Sharjah

- Abu Dhabi to Dubai
- Dubai to Abu Dhabi
- Dubai to Sharjah
- Sharjah to Dubai

Origin/Destination Matrix for UAE administrative regions

2nd pilot (2019-2021)

Pilot countries

- Brazil
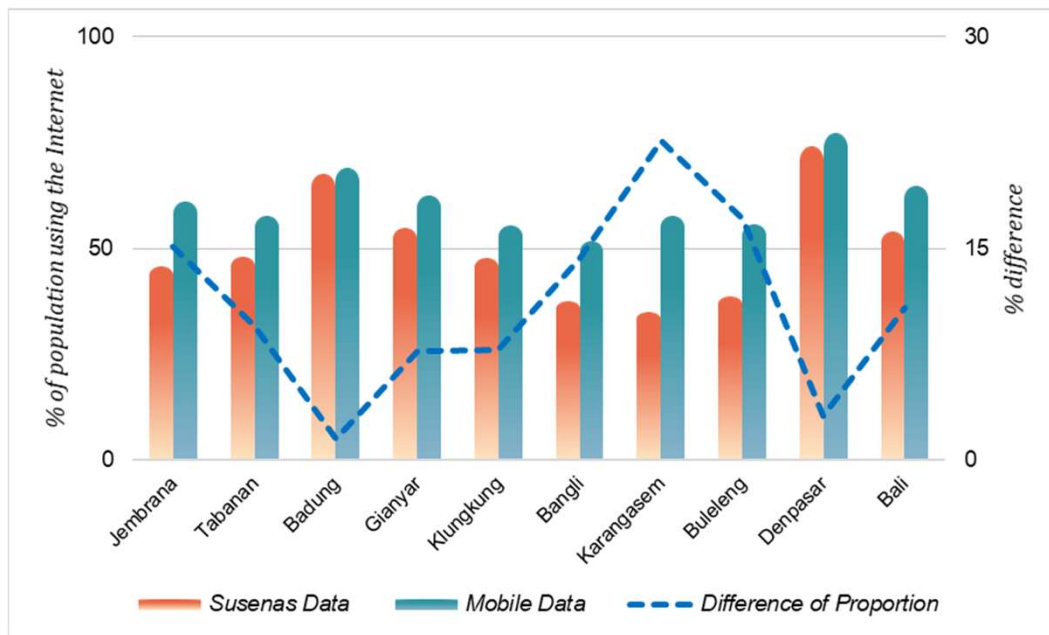- Indonesia

Indicators: 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)

Objectives (2019-2021)

- Develop a methodology (including model(s), algorithm) that can be used for information society indicators included in the SDG monitoring framework

- Develop quality assurance/checks that can be used for mobile phone data for information society statistics.

- Examine how mobile phone big data can be used to estimate ICT access and use of the whole country/population.

- Showcase best practices that can be replicated in other countries.

Results: Percentage of population using the Internet (SDG 17.8.1)
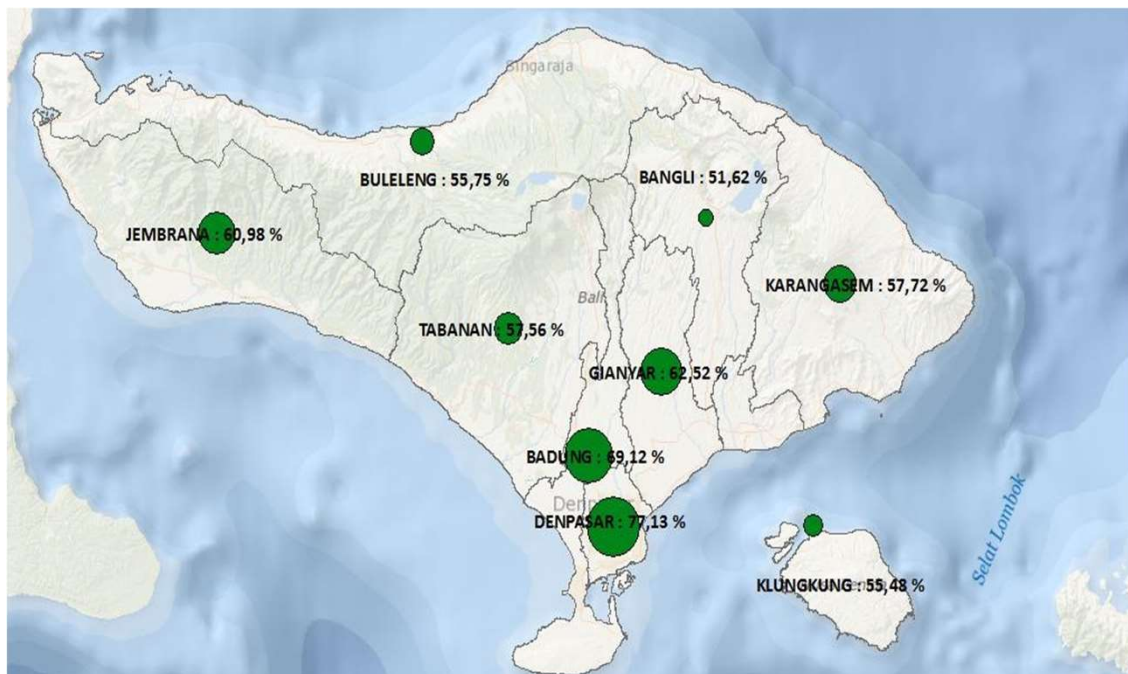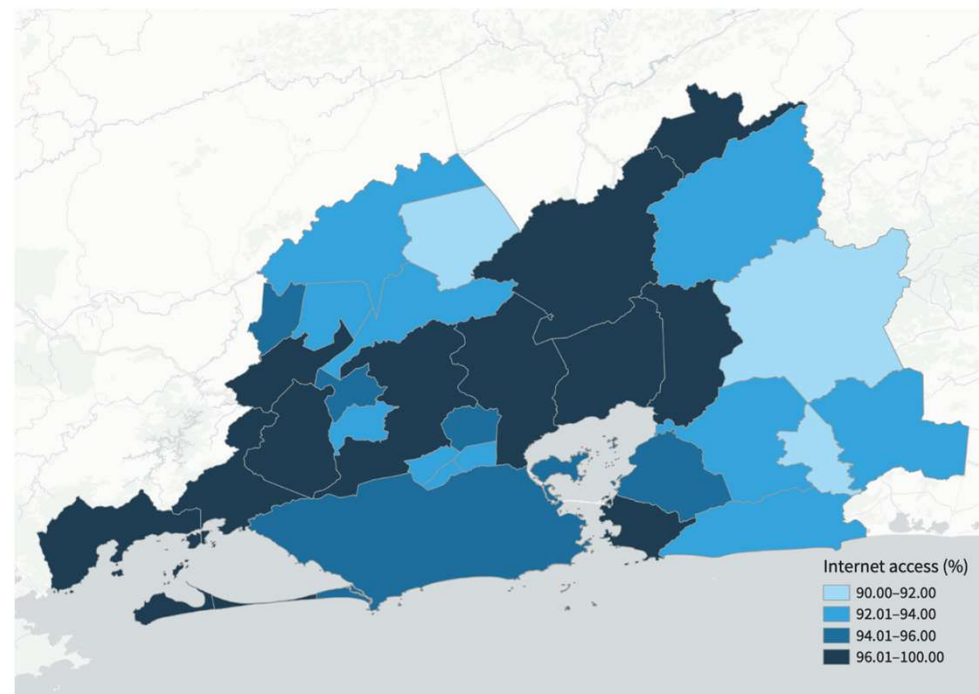
Indonesia
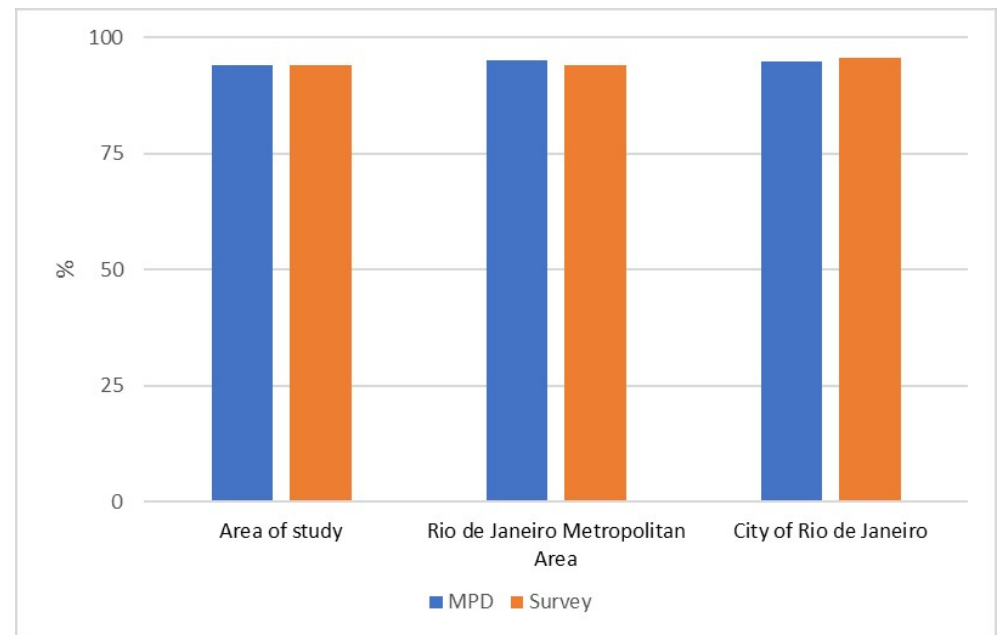
Brazil

Results: Percentage of population using the Internet (SDG 17.8.1)
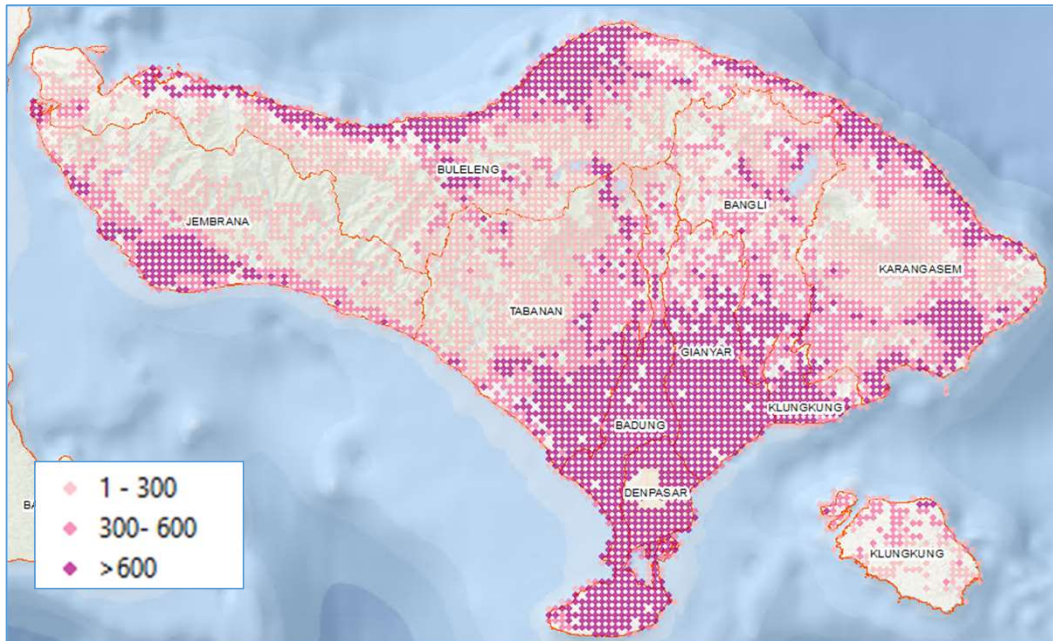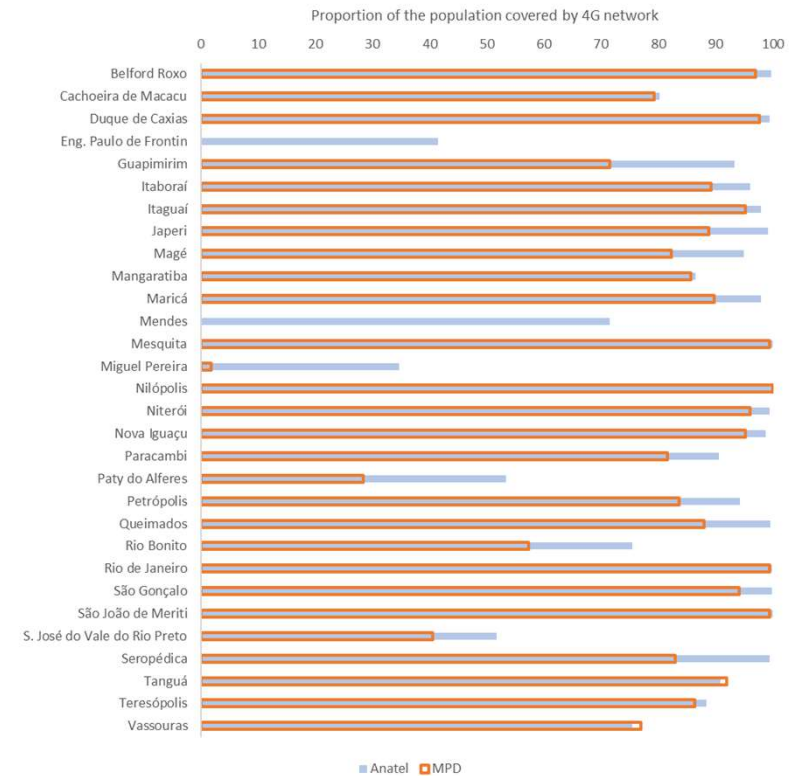
Bali, Indonesia

Brazil

Results: Percentage of population covered by mobile signal (SDG 9.c.1)

Population covered by mobile cellular network (3G) in Bali, Indonesia

Legend:
- 1 - 300
- 300 - 600
- >600

Brazil

Proportion of the population covered by 4G network

Municipalities (top to bottom): Belford Roxo, Cachoeira de Macacu, Duque de Caxias, Eng. Paulo de Frontin, Guapimirim, Itaboraí, Itaguaí, Japeri, Magé, Mangaratiba, Maricá, Mendes, Mesquita, Miguel Pereira, Nilópolis, Niterói, Nova Iguaçu, Paracambi, Paty do Alferes, Petrópolis, Queimados, Rio Bonito, Rio de Janeiro, São Gonçalo, São João de Meriti, S. José do Vale do Rio Preto, Seropédica, Tanguá, Teresópolis, Vassouras

Legend: ■ Anatel □ MPD

Results: Population covered by mobile network (SDG 17.8.1)

Indonesia

Brazil

Proportion of population covered with 3G mobile network
Coverage area radius set at 10km, flat approach

Coverage
- 0
- 1–10%
- 11–20%
- 21–30%
- 31–40%
- 41–50%
- 51–60%
- 60–70%
- 71–80%
- 81–90%
- > 90%

OpenCelliD, WorldPop, Carto

Source: IBGE Municipality Boundaries 2019.

Legend
Main technology (more frequent) to access the internet

2G   3G   4G
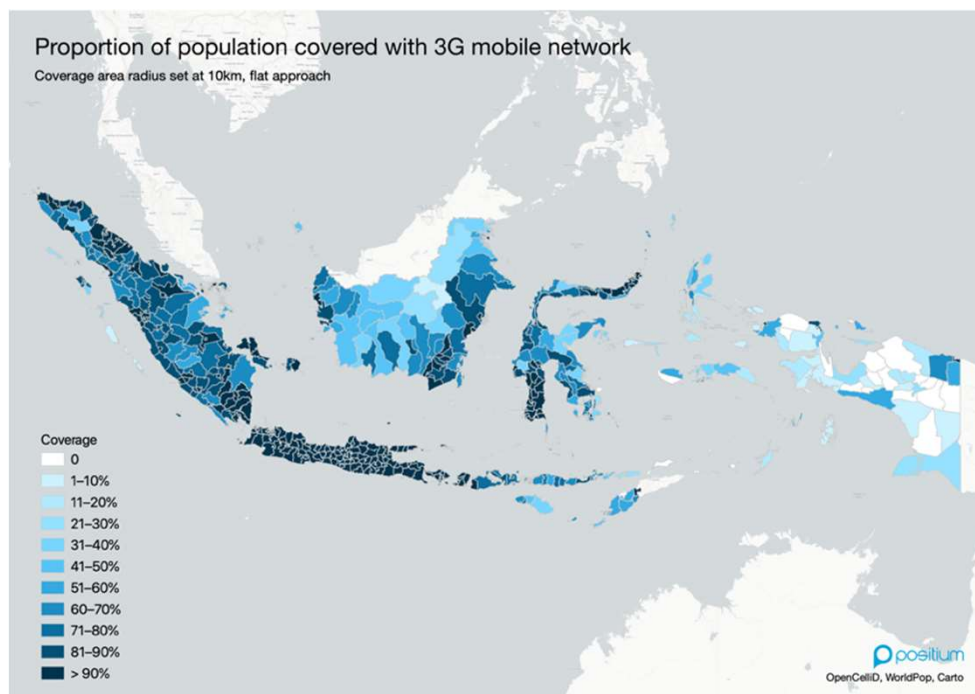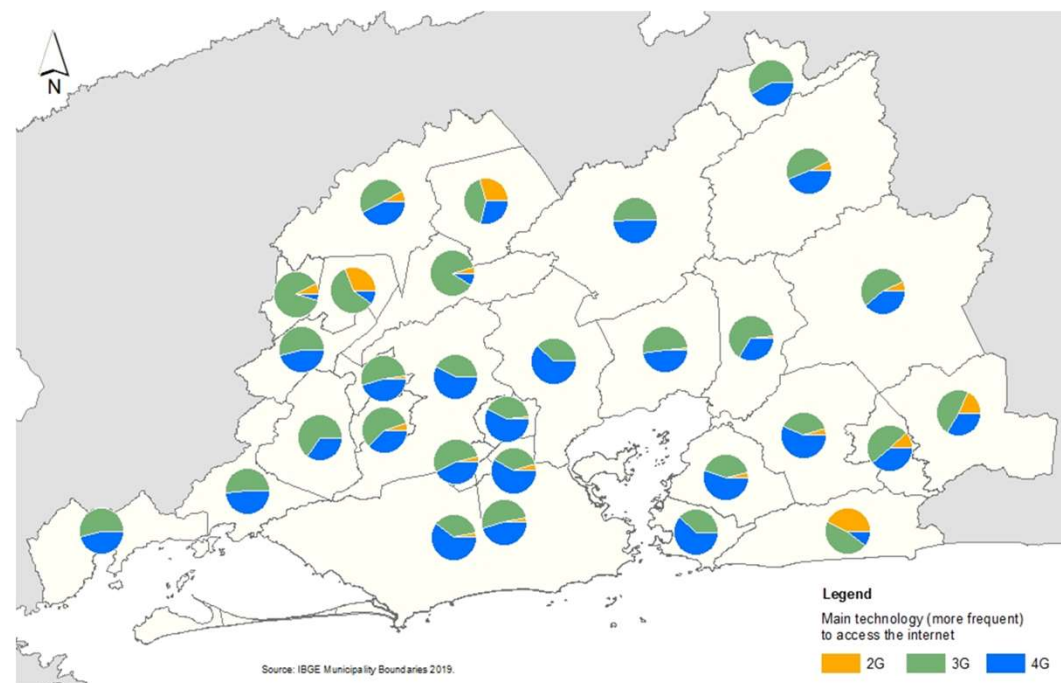
Lessons learned:

MOBILE POSITIONING DATA (MPD) CASE STUDY

Using Mobile Phone Data For Measuring SDGs

Over the past decade, national statistical offices have been tasked with producing accurate, high quality statistics faster, more frequently and with a reduced burden on respondents. Yet they are expected to achieve this within the same or even a smaller budget. To face this challenge, national statistical offices can turn to innovative data sources like passive mobile positioning data (MPD).

Even though the private sector was the first to harness big data, the practice has now expanded to the global statistical community. The United Nations Statistical Commission and national statistical offices are looking into ways of using big data sources to complement official statistics and better meet their objectives for providing timely and accurate evidence for policy-making.

A project, led by ITU, was conducted in 2020 to demonstrate how big data can be used to produce internationally agreed ICT SDG indicators 9.c.1 (Proportion of population covered by a mobile network) and 17.8.1 (Individuals using the Internet). The feasibility of using MPD for both indicators was tested in both Brazil and Indonesia, one of each presented in the case studies. The collaboration showed that public and private sector organisations can work together for societal interest to leave no one behind.

1. National coordination (ministry, regulator, data protection agency, NSO, MNOs)

2. Share country experiences to improve measurements of MPD-ICT indicators

3. Establish data pipelines that can be used for multiple areas of statistics

https://www.itu.int/en/ITU-D/Statistics/Documents/bigdata/ITU_SDG_case_study.pdf

Key takeaways

1. Prepare all administrative and legal procedures to access the data before the project starts
   - Agreed processing model for calculation (by data providers or by TRA/NSO)
   - Agreed method for data transfer

2. Standard, clear and unambiguous methodology:
   - detailed data source description (input data)
   - calculation methodology
   - example algorithms for calculation

3. Infrastructure and human resources (data scientist) for data processing should be in place

4. Coordination with all stakeholders (access to data, validation of results, analyses)

Recommendations

➢ <u>Think of MPD as any new data</u> in a statistical business process

➢ <u>Assess the current situation</u> - data is governed by privacy, telecommunication legislation, and operator internal rules

➢ Coordinate with <u>all stakeholders</u> to avoid surprises

➢ Discuss and agree on <u>all elements</u>, also publication, prior to starting the project

➢ Set <u>realistic timeline</u> – often time-consuming process

➢ <u>"Start small"</u> – sample data helps to design the solution and establish feasibility for statistics

➢ <u>Learn</u> from others - <u>Join</u> the MPD Task Team

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

- Methodological guide on the use of mobile phone data: Measuring the Information Society (SDG ICT indicators): https://unstats.un.org/wiki/display/MPDMIS

- Online training course on mobile phone data

- https://www.itu.int/en/ITU-D/Statistics/Documents/bigdata/ITU_SDG_case_study.pdf

- Handbook on the use of Mobile Phone Data for Official Statistics Methodological Guide and Proposed ICT Indicators Based on Big Data

- 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

Measuring/mapping Internet use at sub-national level

Data on Internet use/access is scarce – especially in Africa

Latest year for data on proportion of individuals using the Internet

- 2018-2020
- 2021-2022

Gather evidence on Internet use/access from countries

Sources

➢ National household surveys (ICTs or other)

➢ National censuses

➢ Business surveys with large sample size and geographical breakdown
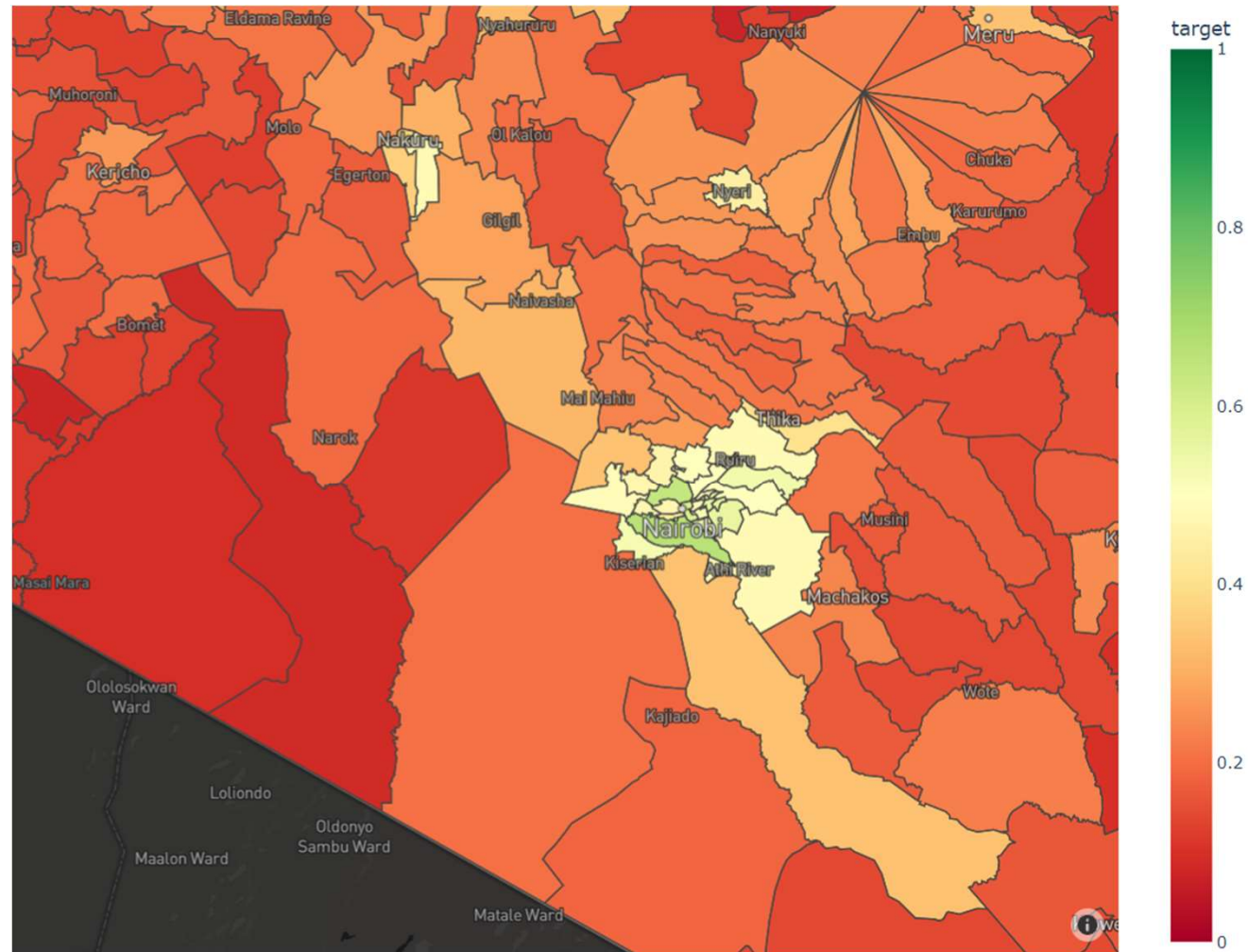
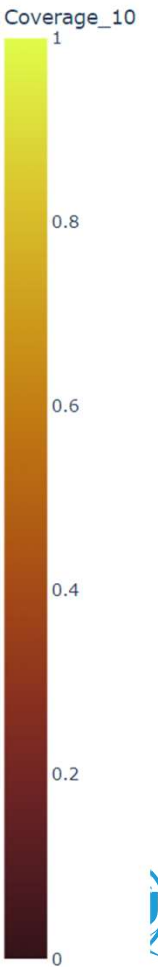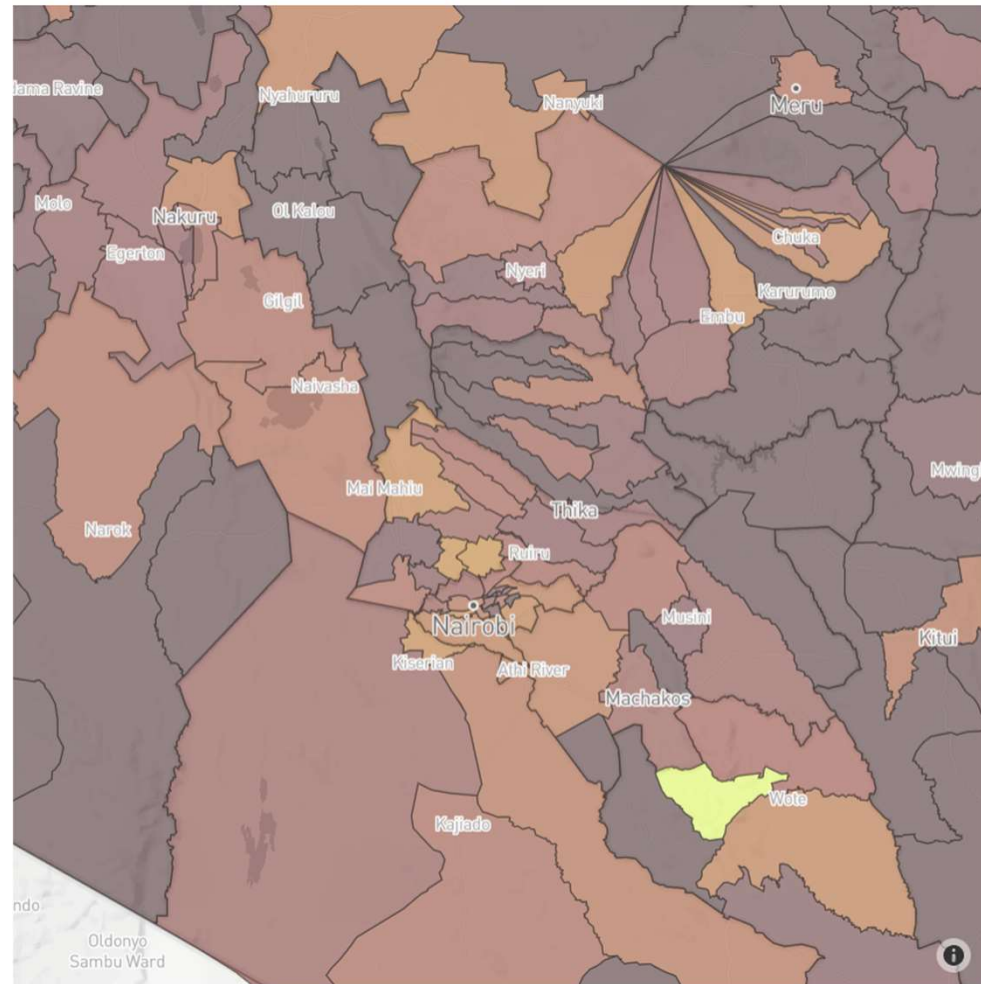# Map global open source data to the same geographic locations

➢ Satellite imagery

➢ Speed measurements

➢ Social media

➢ ICT infrastructure
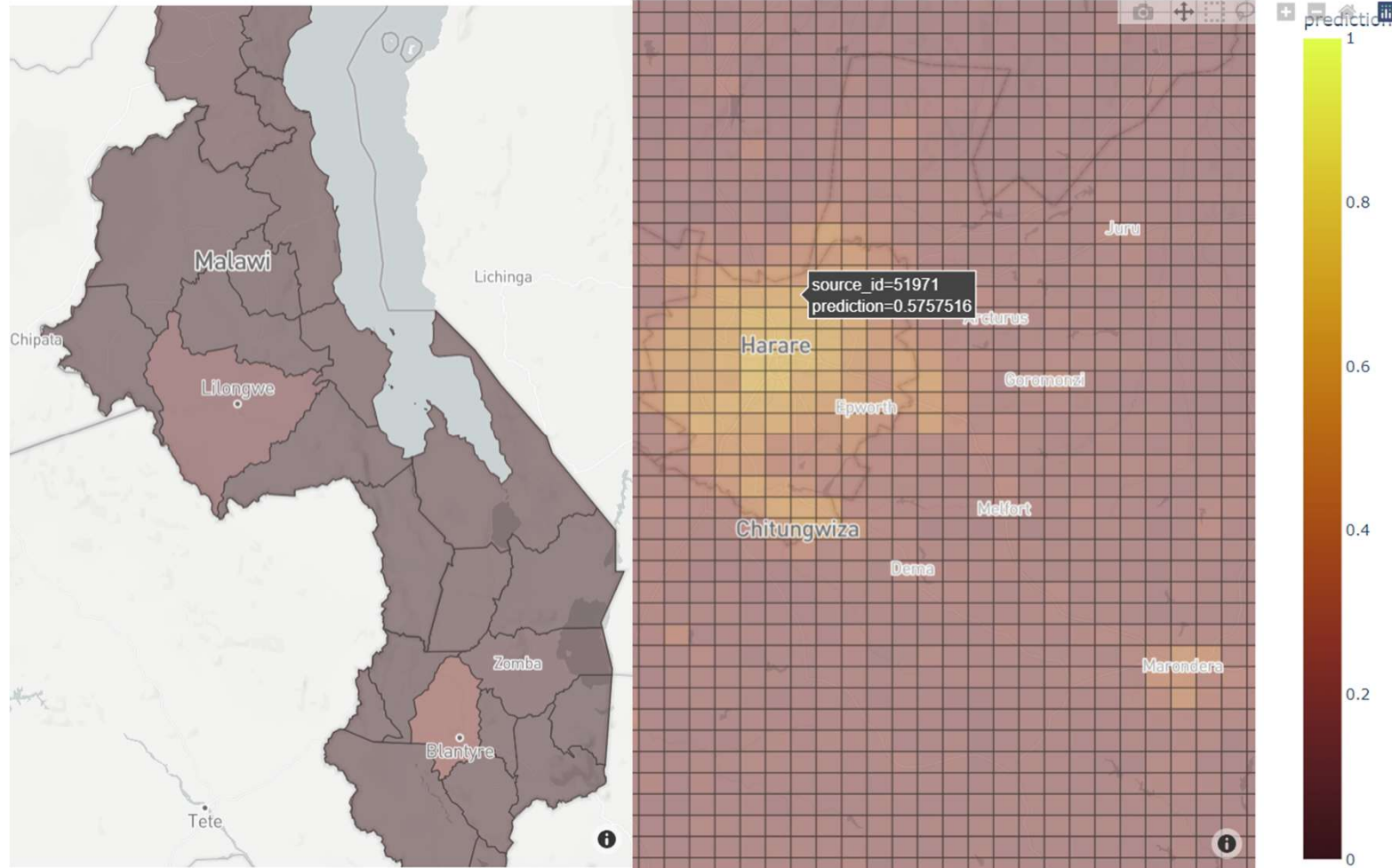
➢ Population and other socio-economic variables

Build statistical model and predict on other countries

Predict at various levels:

➢ Region

➢ Municipality

➢ Grid (local)

Challenges:

➢ Availability of household data

➢ Multiple models may be needed

Thank you for your attention!

✉ indicators@itu.int

For more information, please visit our website:

https://www.itu.int/itu-d/sites/statistics/