**Webinar UN Regional Hub for Africa**

# Leveraging Mobile Phone Big Data for Information Society Statistics

16 May 2024

## Leveraging mobile phone data for policy – an introduction

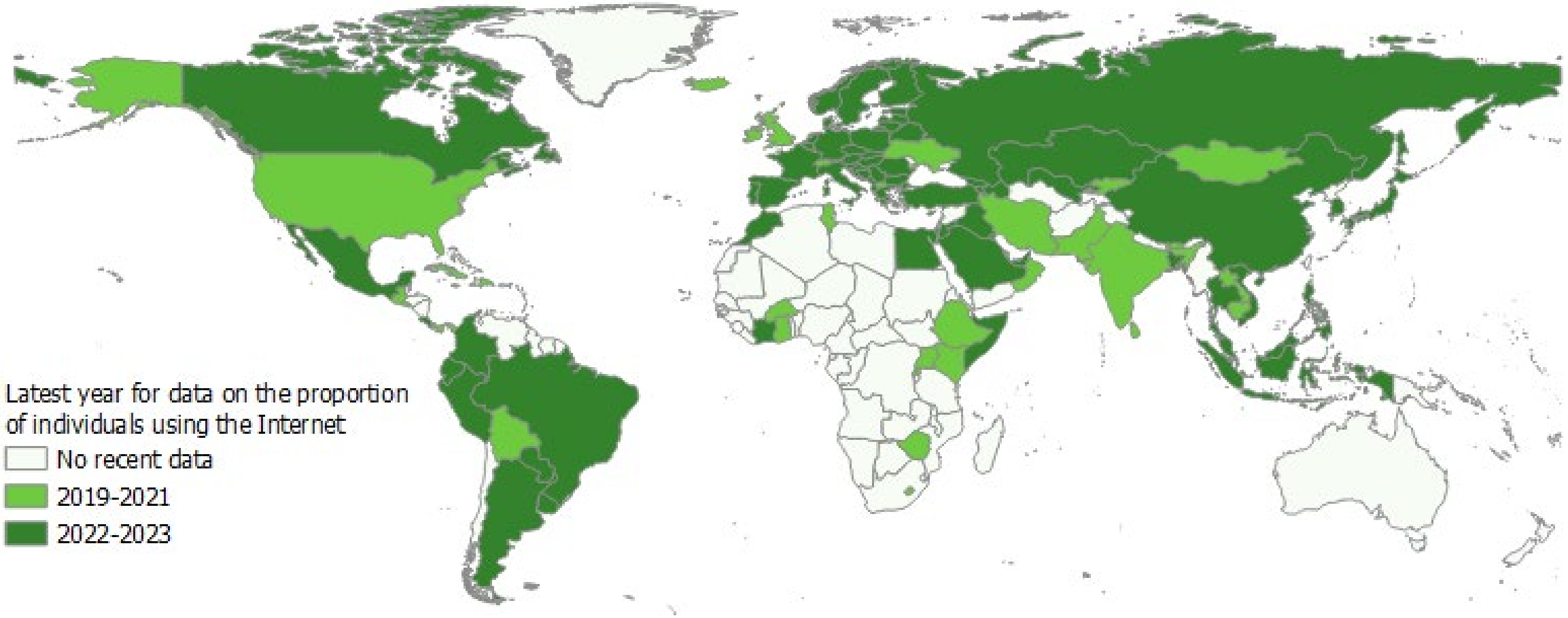
**Fredrik Eriksson**
Data Scientist
International Telecommunication Union (ITU)

# Outline

1. Leveraging mobile phone data for policy
2. Technical environment and processing models
3. Data quality checks and reference data
4. Practical considerations

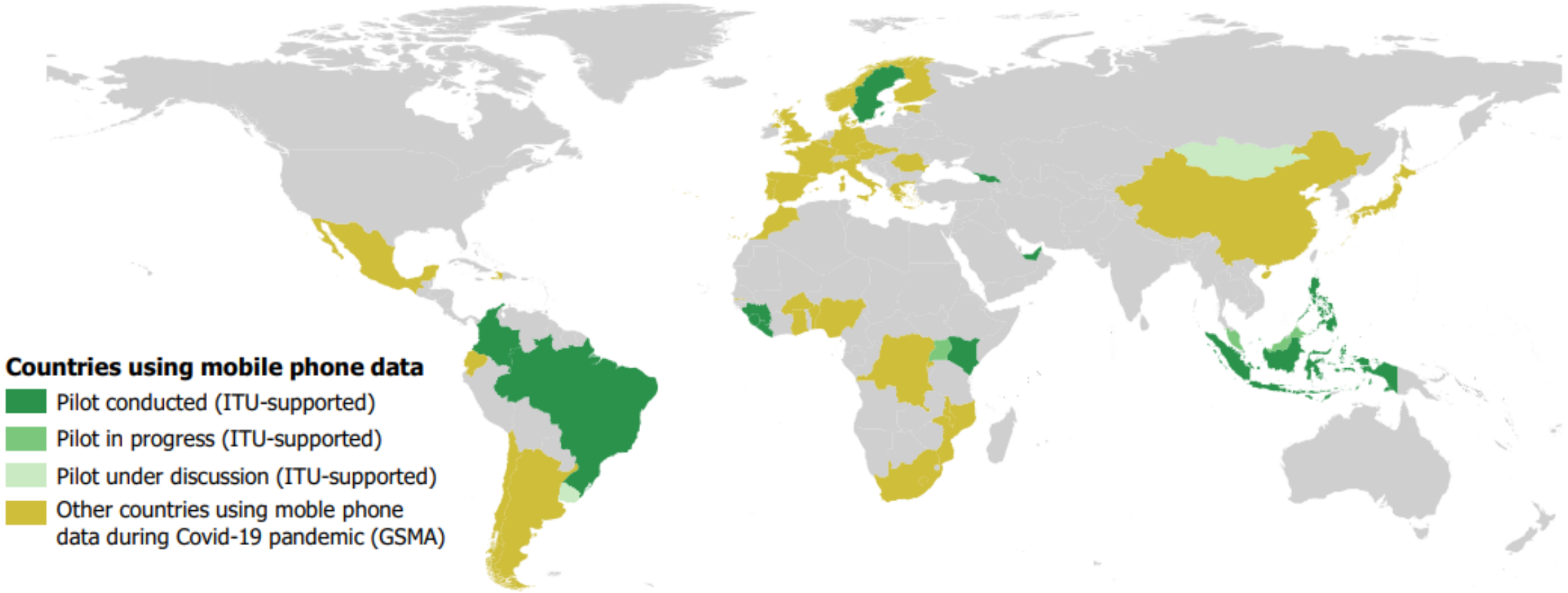
1. Leveraging mobile phone data for policy

# Data on connectivity is scarce – especially in Africa



1. Leveraging mobile phone data for policy

# Mobile phone data is increasingly used across the world




ITU-supported pilots include:
Conducted: Guinea, Liberia and Sierra Leone (2015); Colombia, Georgia, Kenya, Philippines, Sweden and United Arab Emirates (2016-2017), Brazil and Indonesia (2019-2021)
In progress: Malaysia and Uganda
Under discussion: Mongolia and Uruguay

1. Leveraging mobile phone data for policy

# Mobile phone data can be used across a variety of domains

| Policy area | Use |
|---|---|
| ICT | Network development / optimization |
| Transport | Optimize transport routes |
| Tourism | Tourist origin / movement |
| Migration / dynamic population mapping | Commuting / refugees resource planning |
| Disaster preparedness / monitoring | Early warning systems / displacement of people |

Image generated using Dall.E 2 by OpenAI

# What is mobile phone data?

## Call detail records / IP detail records

- Already collected by MNOs for billing
- Limited to events (Calls, SMS, data)
- Actual data includes
  - Time (and duration) of events
  - Cell tower location of events

## Mobile positioning data

- Regular monitoring of data flows between network entities (passive)
- More details but larger data volume

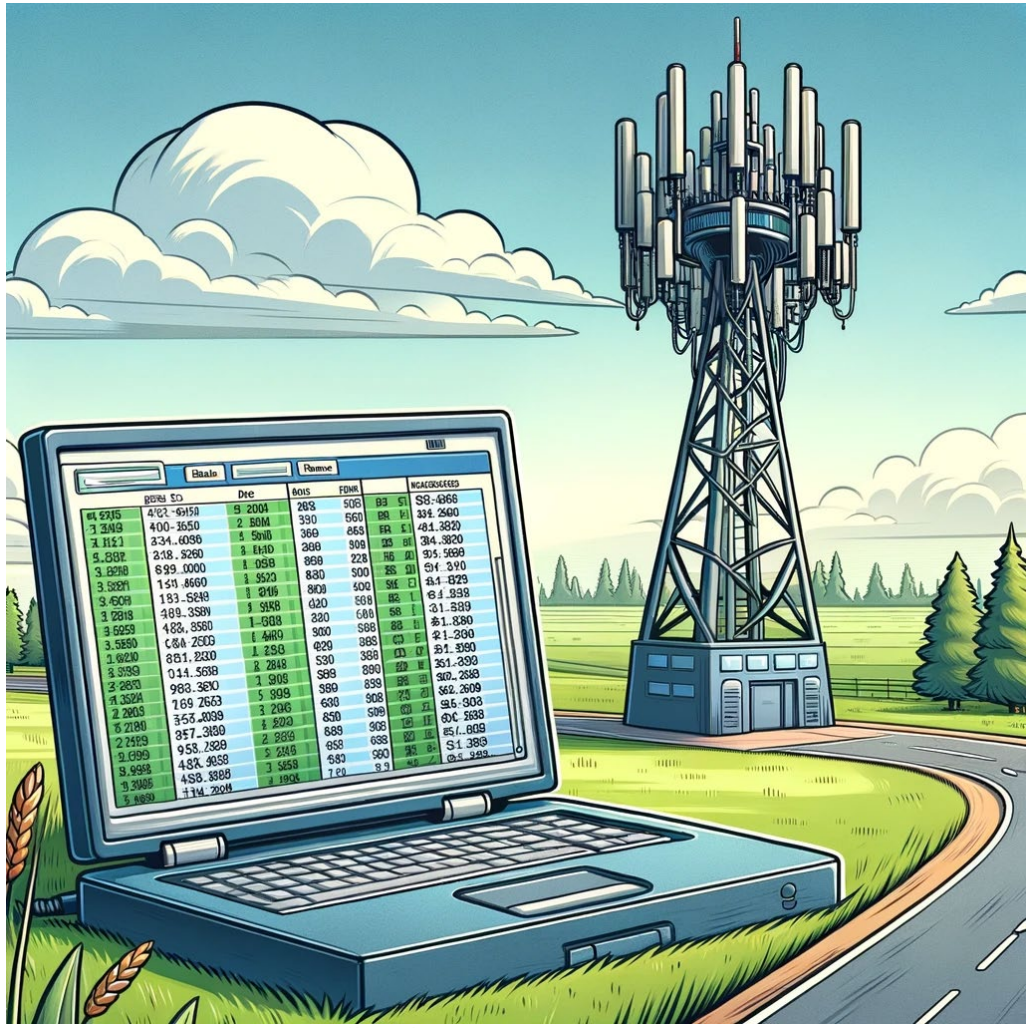


Image generated using Dall.E 2 by OpenAI

2. Technical environment and processing data

# No one-size-fits all for working with mobile phone data

- **Three processing methods**
  - Operator-led
  - Agency-led
  - Public-private partnership

- **Big data infrastructure and skills needed**

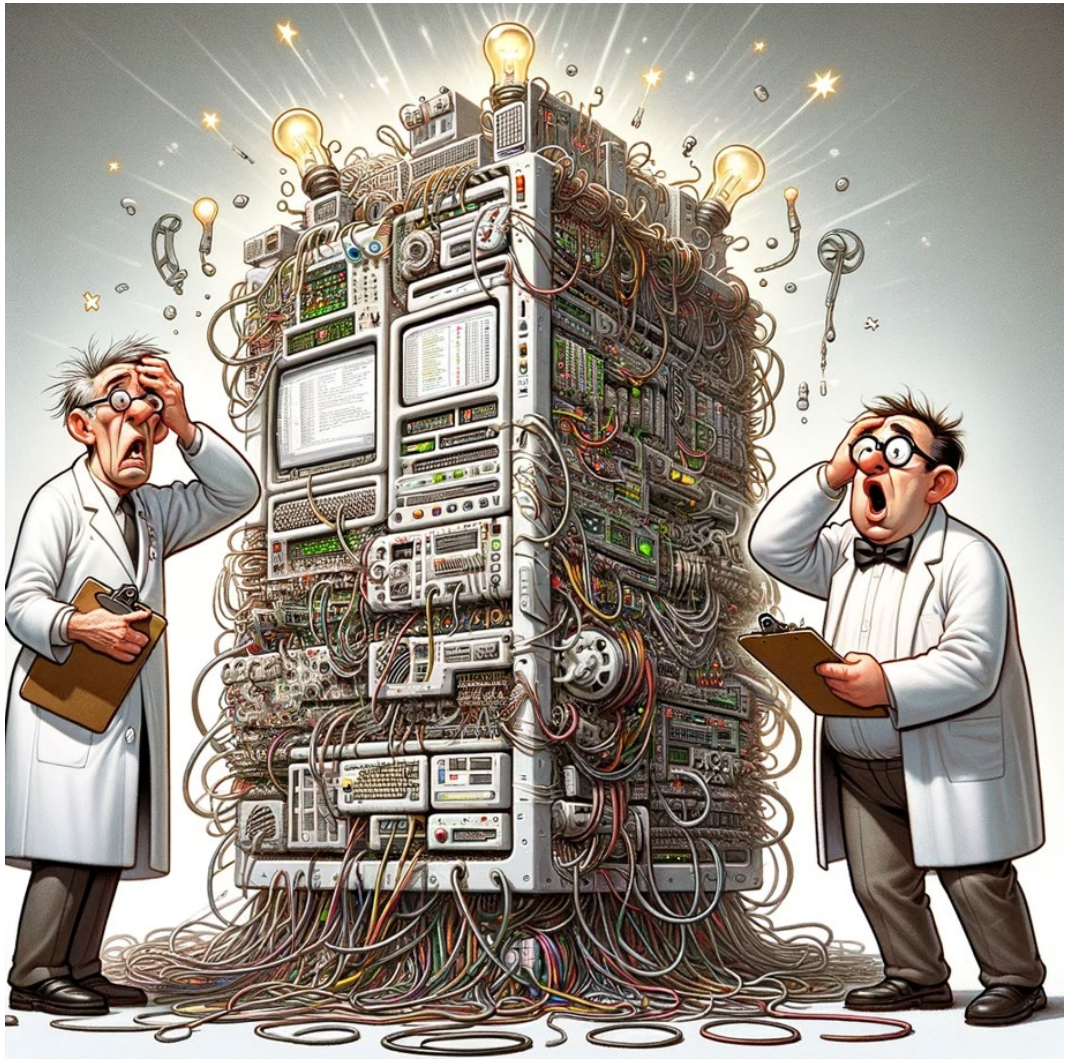
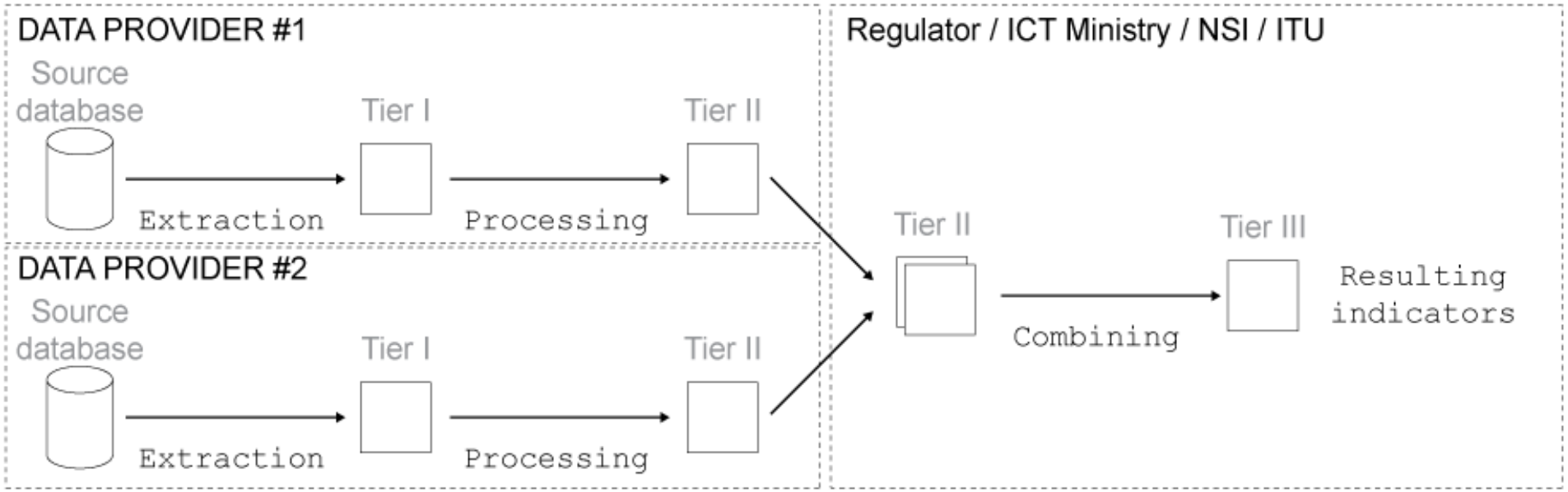


Image generated using Dall.E 2 by OpenAI

# Option 1: Operator-led processing



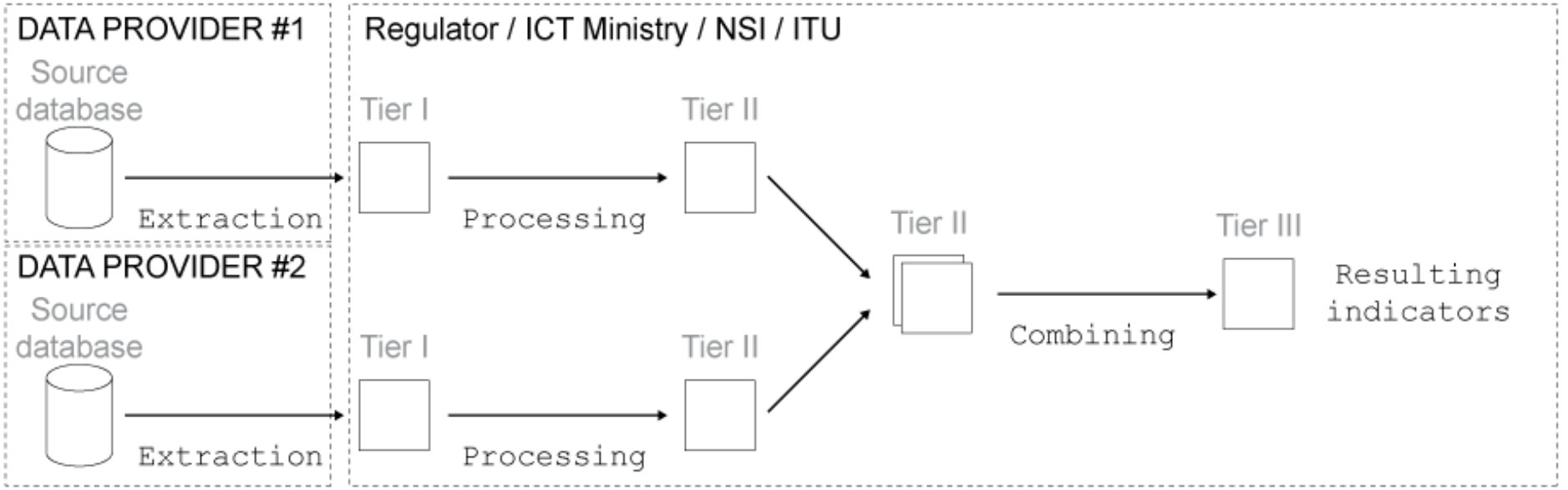

Tier I: Raw or anonymized data

Tier II: aggregated data

Tier III: final indicators

- Higher workload for **operators** - need clear commitment and resources
- **Less control and granularity** - training required for operators

# Option 2: Agency-led processing



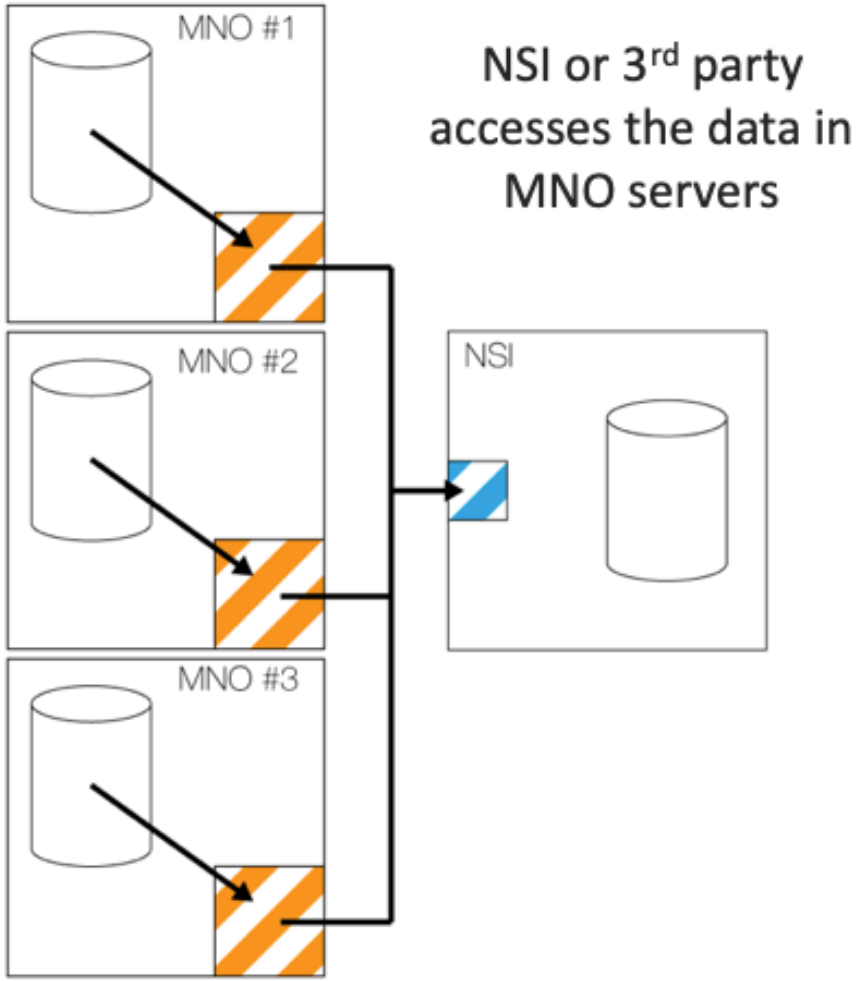

Tier I: Raw or anonymized data

Tier II: aggregated data

Tier III: final indicators

- Higher workload for **regulator / government agency**
- **More control and granularity**, but need own human/technical resources
- **Sensitive data transfer** - more emphasis on privacy protection methods

# Option 3: Public-Private Partnership




- High workload for both operator and agency
- Higher control for both operator and agency
- Greater data security / Lower risk of (accidental) disclosure
- More complex environment and higher cost – suitable for long-term engagement

# Big data technical environment

- Fast evolution from Hadoop to Spark -> myriad of options
- On-premise vs. cloud?
- Factors to consider:
  - Current technical environment
  - Technical skills / IT support
  - Financial resource
  - Security / control
  - Legal restrictions / requirements
  - Real-time needs vs batch

# Data processing and quality checks

- Data pre-processing and merging with other datasets, e.g. CRM
- Anonymization / pseudonymization (done by operator)
- Data aggregation
- Data quality checks on:
  - Duplicated / missing values
  - Outliers
  - Events / Activity over time
  - Events per subscriber (“robots?”)
  - Realistic cell location activity?

Image generated using Dall.E 2 by OpenAI

# Reference data is key to turn mobile phone data into indicators

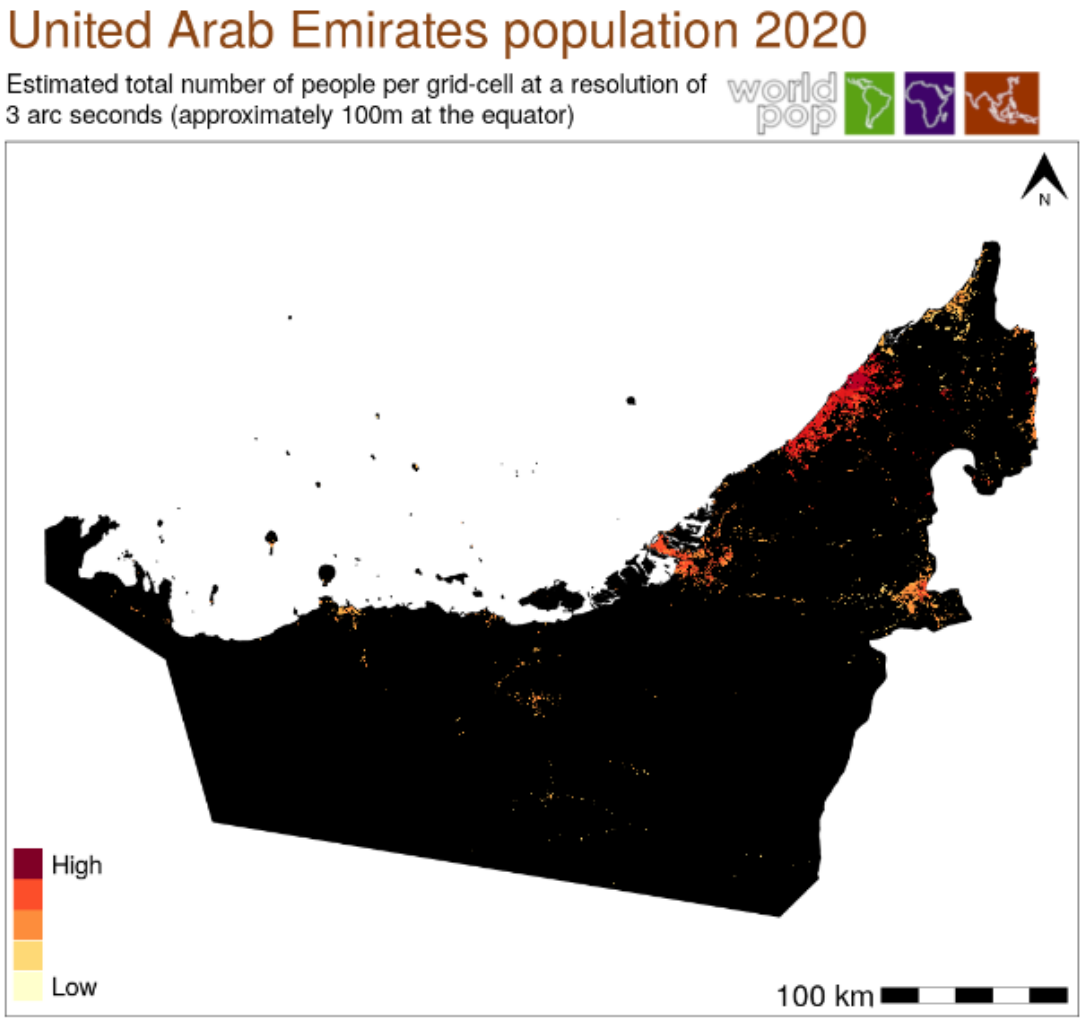
- Local Administrative Units (geodata of administrative borders)
- Granular population estimates
- Location of cell towers
- Geographic elevation
- Household or other official data (for validation)

Image generated using Dall.E 2 by OpenAI

# Example: Detailed population data

Granular population estimates can come from:

- National Statistics Offices
- WorldPop
- High Density Population Maps (Meta)




WorldPop (worldpop.org - School of Geography and Environmental Science, University of Southampton
©2020 This work is licensed under a Creative Commons Attribution 4.0 International License

## 4. Practical consideration

# Tips of how to get started ?

- Think of MPD as any new data
- Assess the current situation
  - Legislation
  - Business models
  - Available resources (human, technical financial)
- Contact / invite all stakeholders
  - Discuss / agree on all elements of the project
- Set realistic timeline
- Start with sample data

Image generated using Dall.E 2 by OpenAI

# Additional resources

- UN Big data task team on mobile phone data
- ITU Big Data pilots
- FlowGeek knowledge center
- Online training course on mobile phone data
- Five principles for maintaining public trust

Or contact us at: indicators@itu.int

Image generated using Dall.E 2 by OpenAI

# Thank you!

indicators@itu.int