Big Data for Measuring the Information Society

INTERNATIONAL TELECOMMUNICATION UNION

BIG DATA PROJECT - INNOVATIVE WAYS TO UTILIZE BIG DATA AS A NEW DATA SOURCE FOR ICT INDICATORS

MARGUS TIRU. PROJECT COORDINATOR, ITU CONSULTANT

The Objective

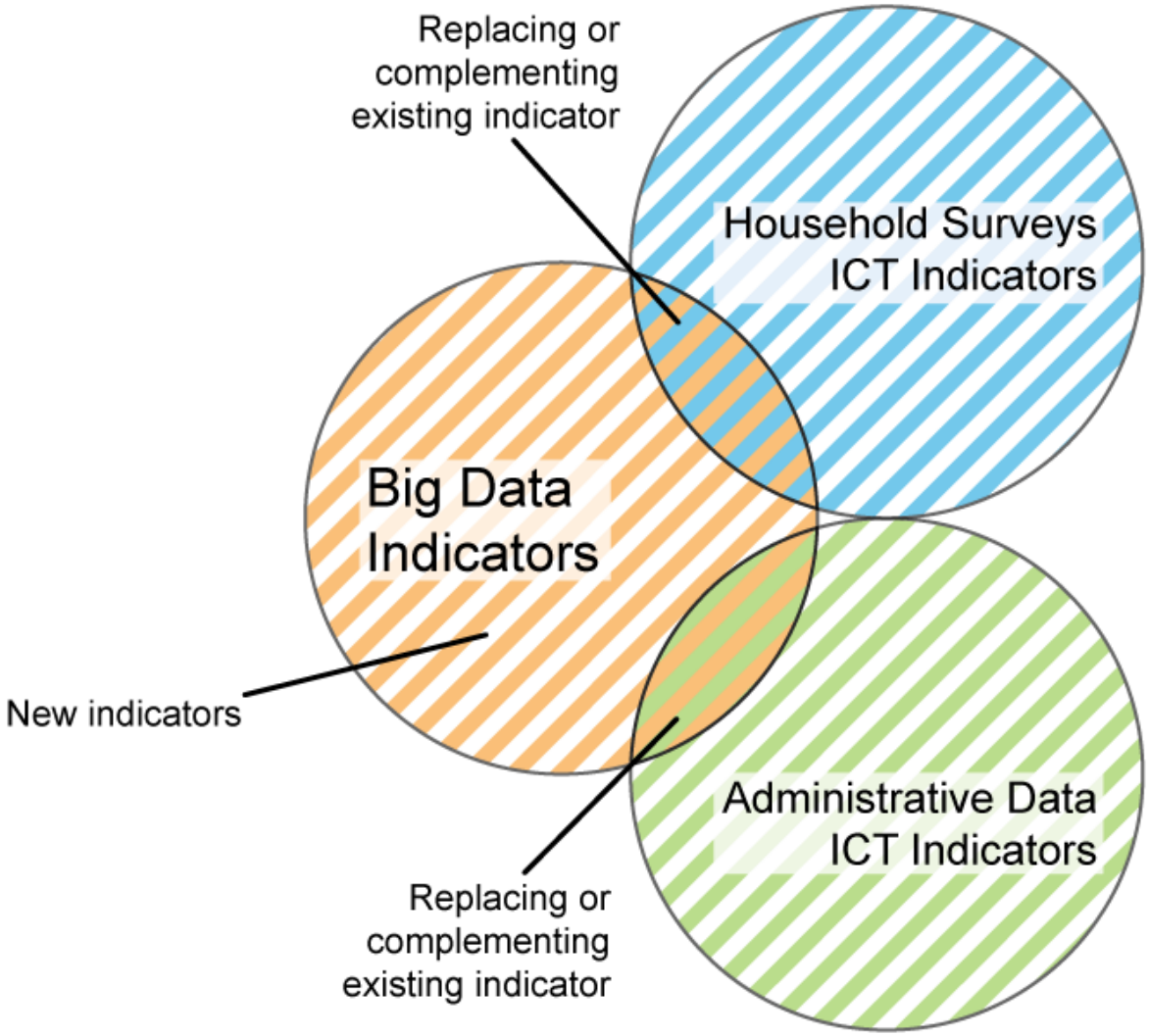
Demonstrate how Big Data can be used for ICT measurement – to produce new and existing ICT indicators to enhance data availability, benchmarks and methodologies to measure the information society

Big Data Indicators

Replace existing indicators

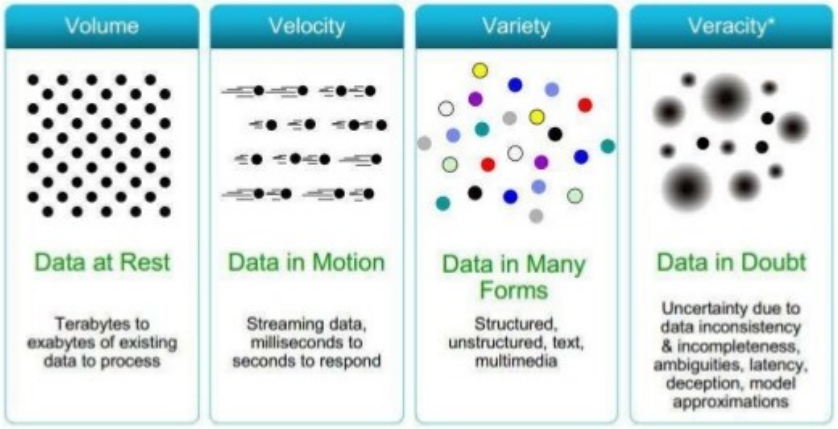
Complement existing indicators (granularity, disaggregation)

New indicators

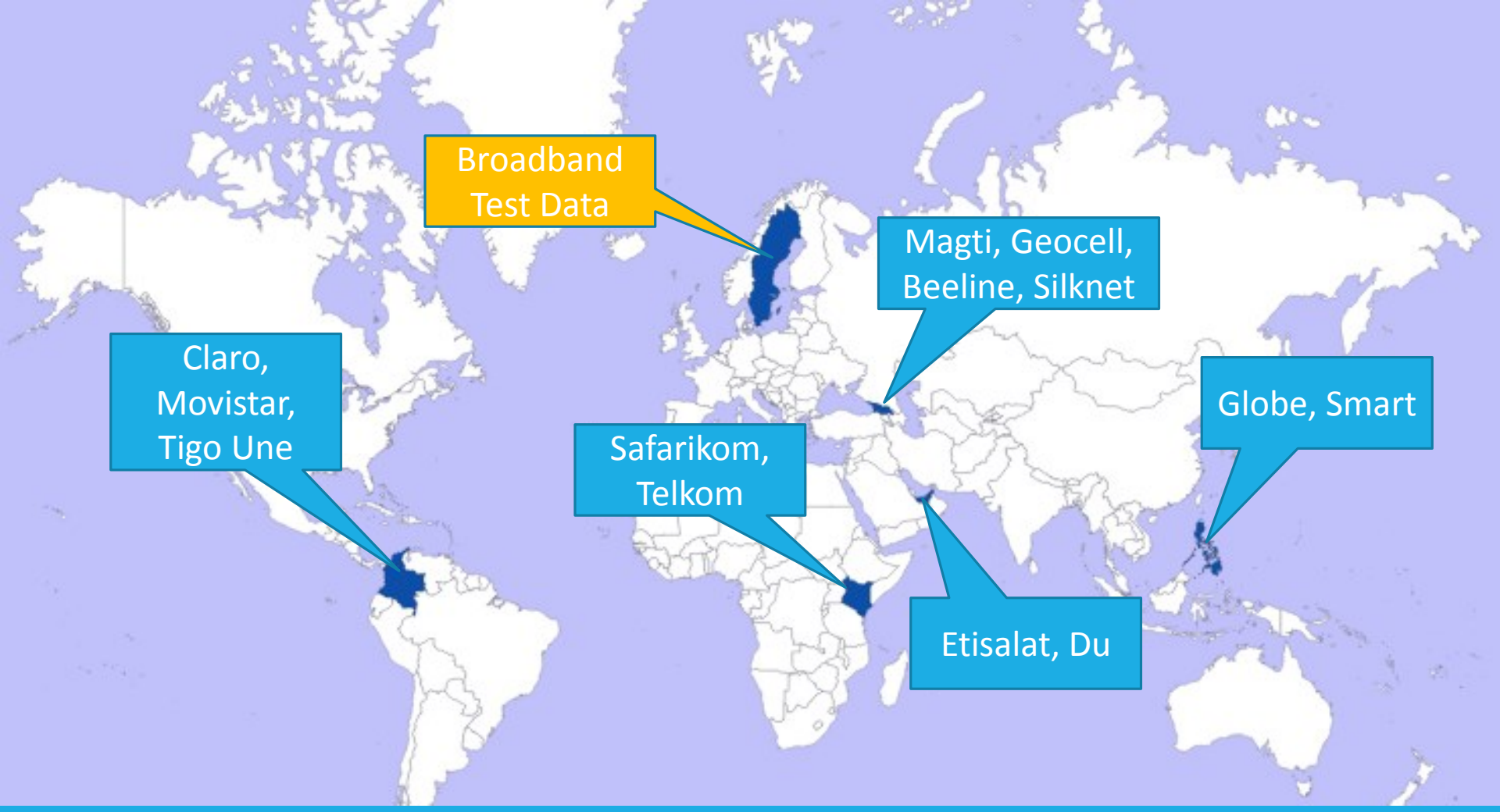


Big Data

In this project we consider Big Data sources from **Mobile Network Operators (MNO)** and **Internet Service Providers (ISP)**, who possess valuable residual data from the aspect of ICT



Source: http://well-managed-business-intelligence.blogspot.in/2012/06/big-data-fourth.html



6 Pilot Countries

The Philippines, Georgia, Colombia, Kenya, UAE, Sweden

Local Stakeholders

Telecommunication regulator / ICT Ministry

National Statistical Institutes

Data protection commission/agency

MNOs and ISPs – data providers:

- Globe, Smart, Magti, Geocell, Beeline, Silknet, Claro, Movistar, Tigo Une, Safarikom, Telkom, Etisalat, Du

Pilot Studies' Challenges

Access to the data

- Legal clearance (regulations)
- Administrative aspects
- Data protection (DPA)
- What source data are collected and available

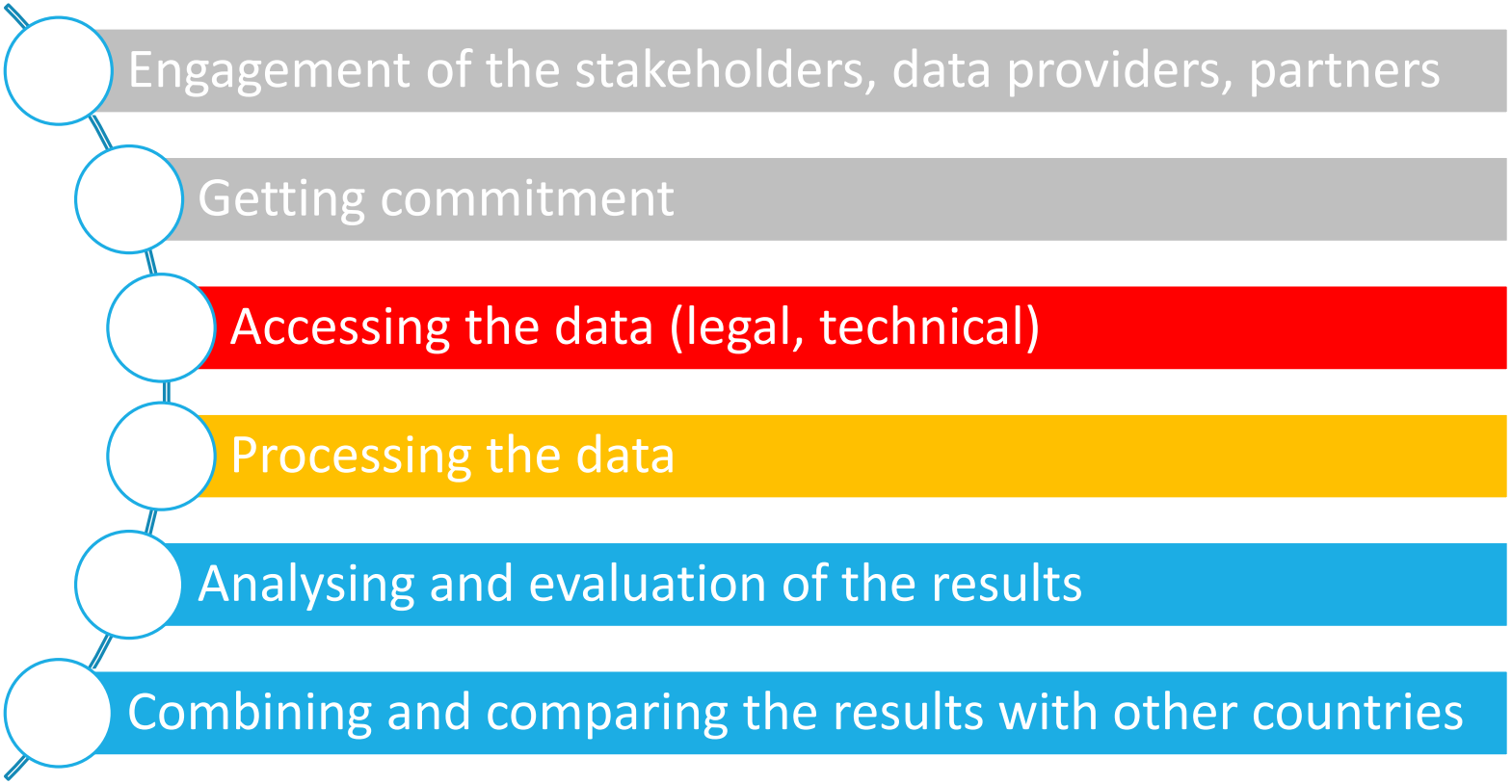
Processing and analysis of large data sets

- Location of the processing
- Methodology for processing, quality of the data and indicators

What indicators can be calculated?

- Are the resulting indicators valuable and usable foe policy and investments decisions?
- Is the data comparable nationally, internationally, over the time?

Status


Engagement of the stakeholders, data providers, partners
Getting commitment
Accessing the data (legal, technical)
Processing the data
Analysing and evaluation of the results
Combining and comparing the results with other countries

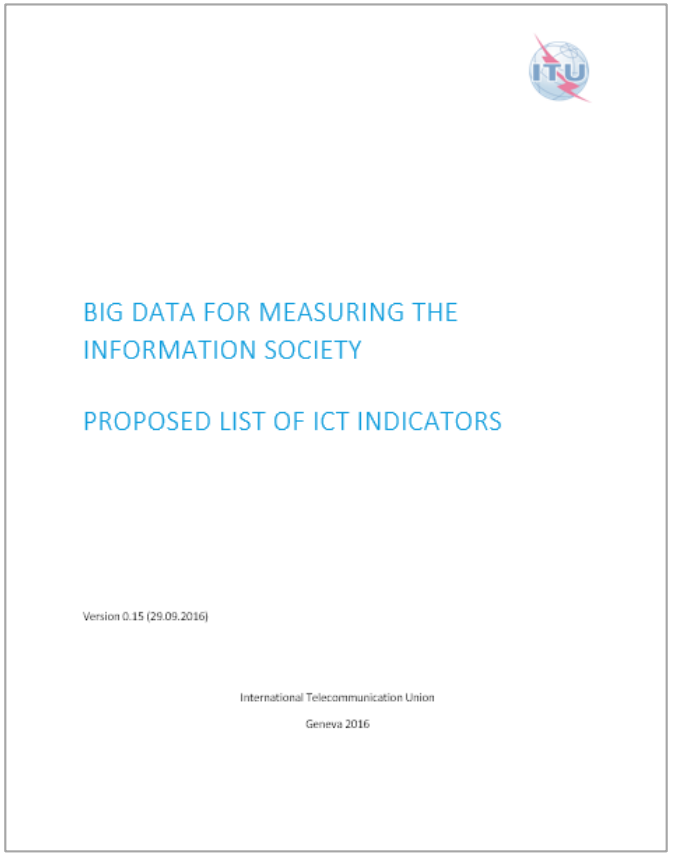
Country	Status overview	Location of the processing	Expected data scientist's visit	Country report ready
Sweden	Needs resolving contract issues and privacy obfuscation	ITU	December	January
Philippines	Waiting for legal go from data providers	Data providers	January	February
Georgia	Data request from data providers, preparations from data providers to send data, GNCC to process data	GNCC	February	March
Colombia	Providers committed, but lack of resources and time now, 2017 Q2 expected	Mixed	March-April	April
Kenya	Ready, need to resolve some aspects of data processing and administrative questions	Data providers	January	February
UAE	Official request made from TRA, data providers preparing for processing of the data	Data providers	January	February

Methodology

Methodology document

- Proposed indicators
- Data source description
- Processing methodology
- Disaggregation
- The purpose and value of the indicator

Draft document to be complemented and amended during the project


BIG DATA FOR MEASURING THE INFORMATION SOCIETY

PROPOSED LIST OF ICT INDICATORS

Version 0.15 (29.09.2016)

International Telecommunication Union

Geneva 2016

Proposed Big Data Indicators

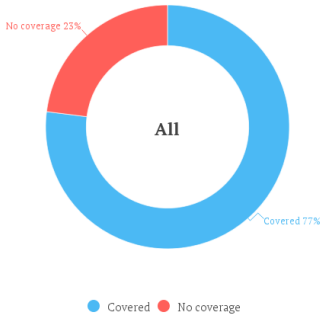
- BD01: Percentage of the Land Area C
- BD02: Percentage of the Population C
- BD03: Usage of Mobile-Cellular Netw
- BD04: Usage of Mobile-Cellular Netw
- BD05: Number of Subscriptions with
- BD06: Active Mobile Voice and Broad
- BD07: Average Number of Active Mo
- BD08: Active Mobile Devices
- BD09: IMEI Conversion Rate
- BD10: Fixed Domestic Broadband Tra
- BD11: Mobile Domestic Broadband T
- BD12: Mobile International Broadbar
- BD13: Inbound Roaming Subscriptior
- BD14: Fixed Broadband Subscriptions
- BD15: Fixed Broadband Subscriptions
- BD16+: any proposed indicators from

Example results of the proposed indicators

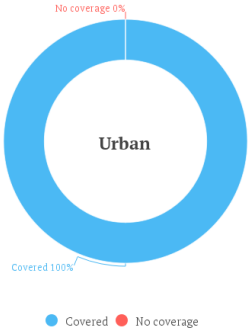
BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

All Technologies

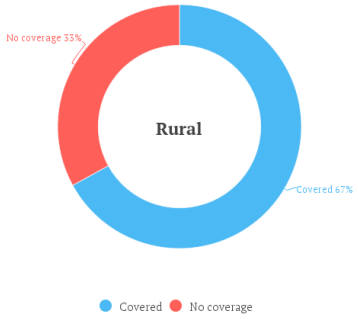
All



Urban



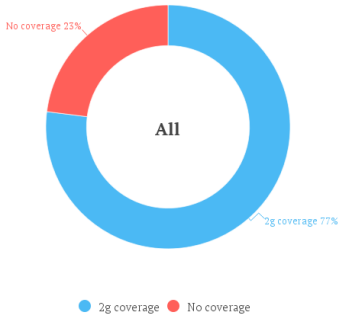
Rural



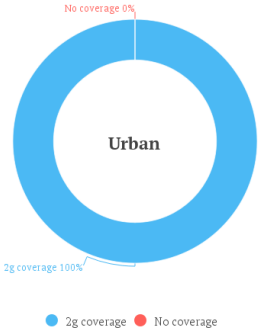
BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

2G

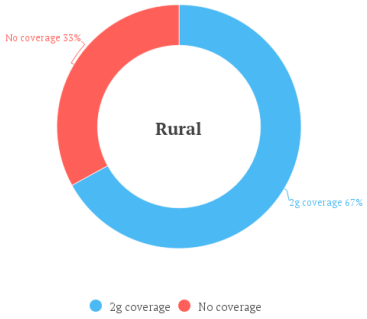
All



Urban



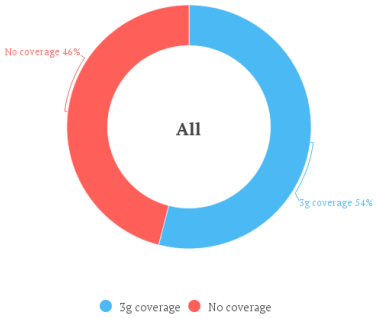
Rural



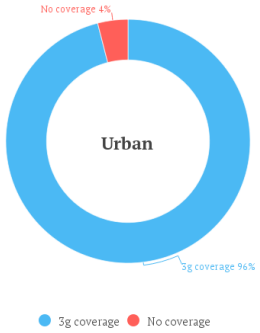
BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

3G

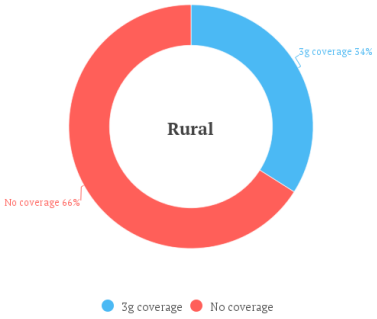
All



Urban



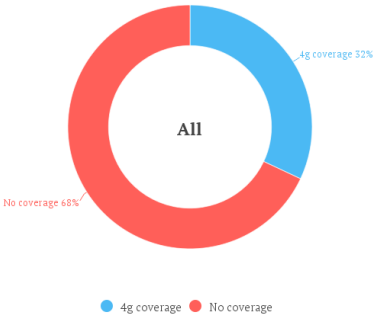
Rural



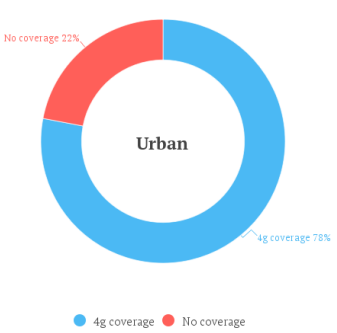
BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

4G

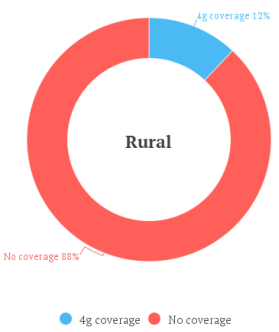
All



Urban



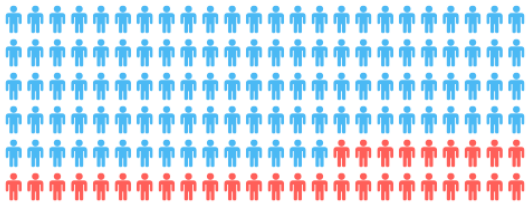
Rural



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

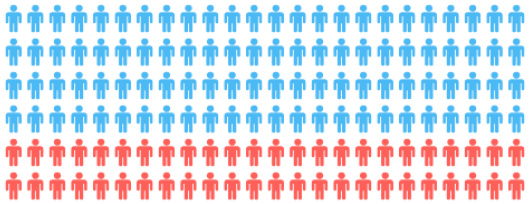
All Technologies

All



Urban

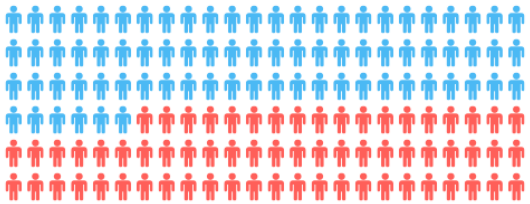
Rural

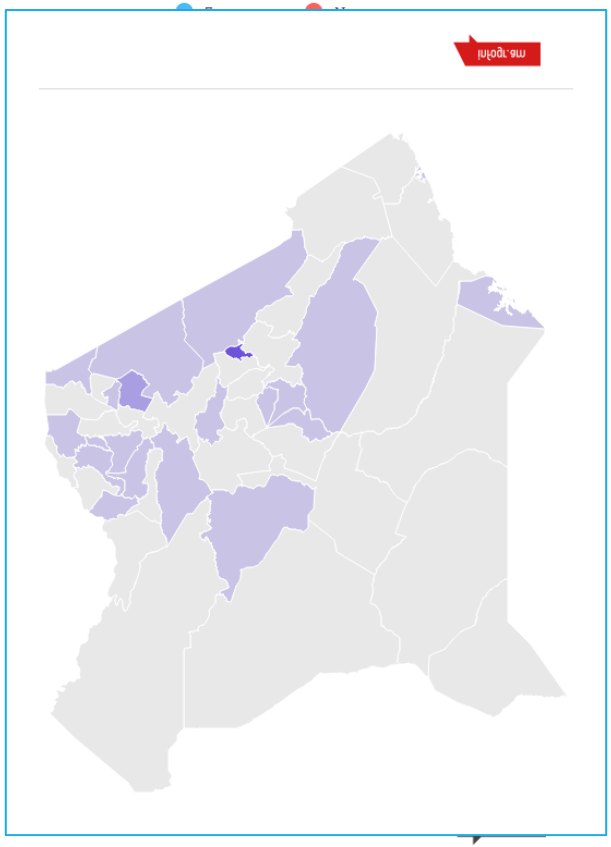


BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

3G

All

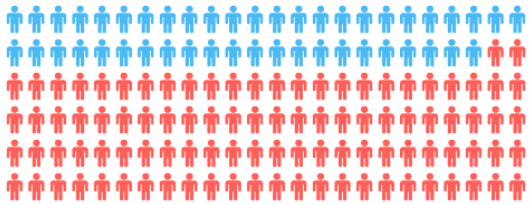


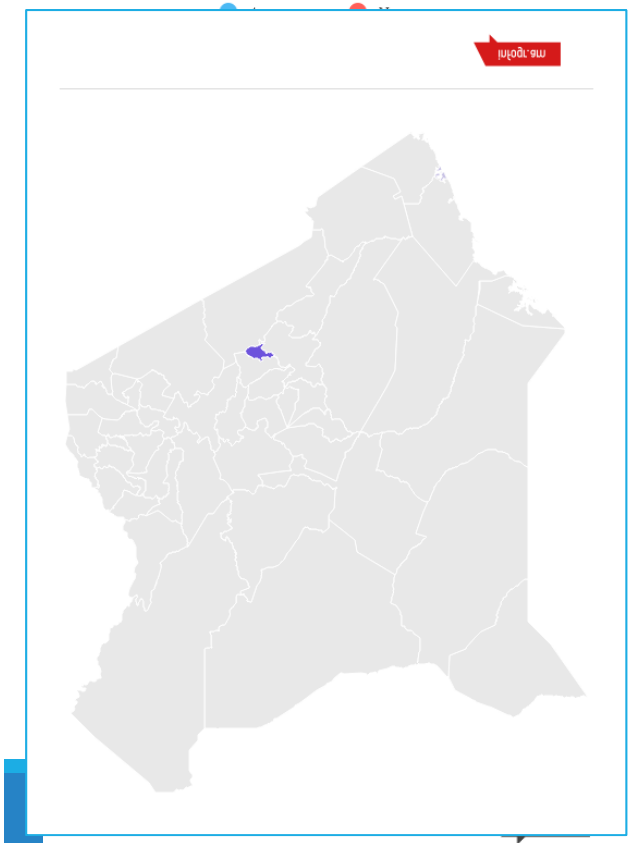


BD01: Percentage of the Land Area Covered by Mobile-Cellular Network, by Technology

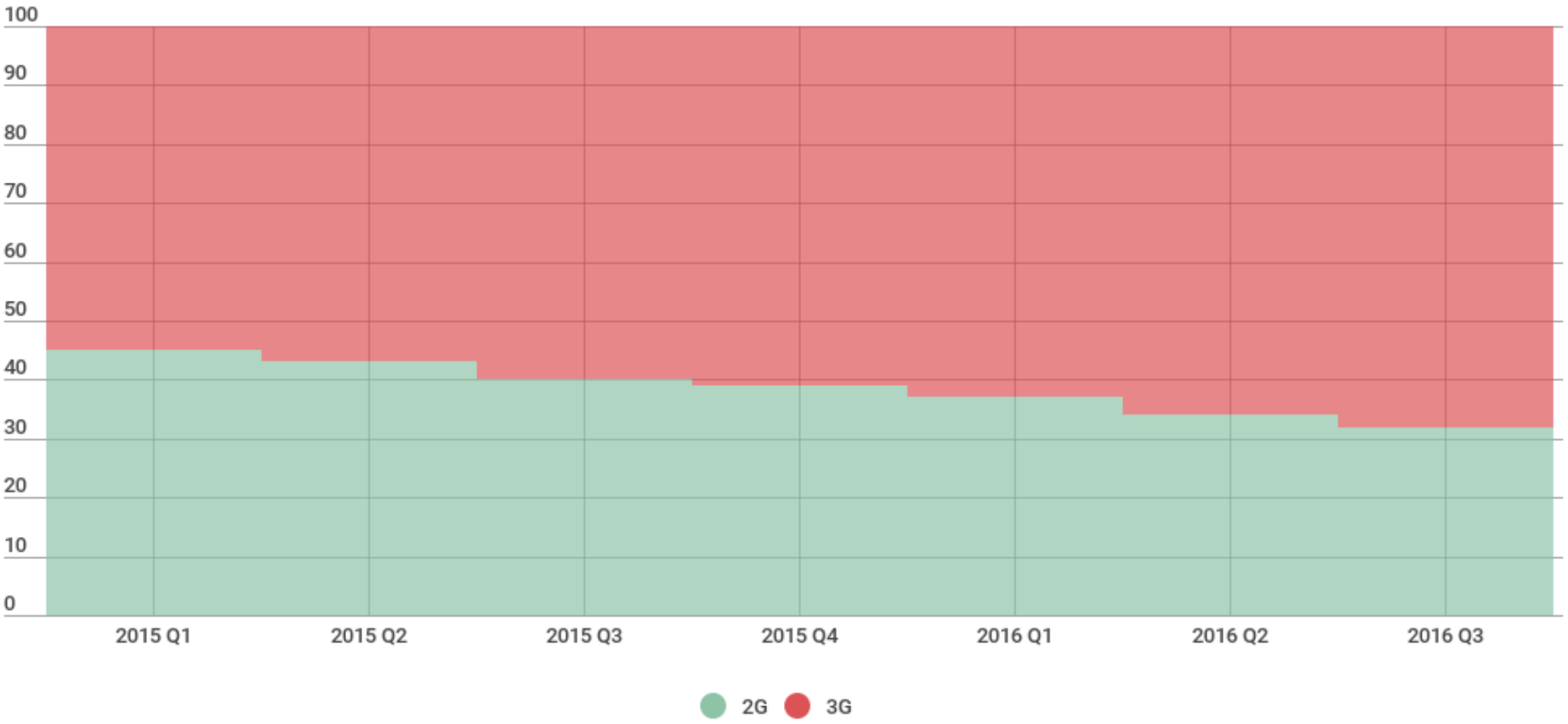
4G

All

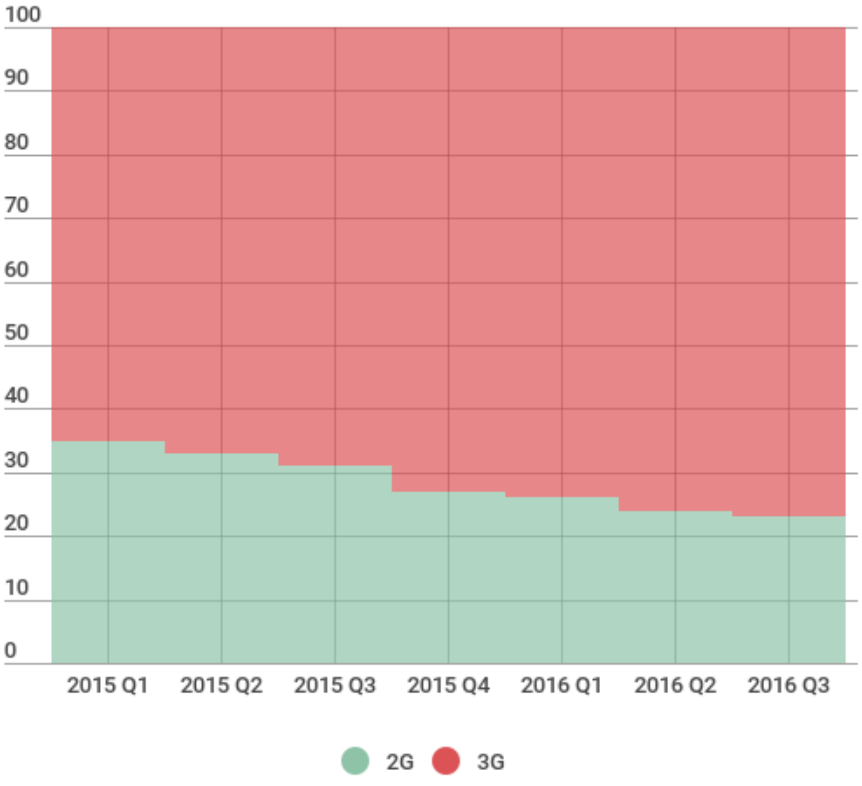




BD03: Usage of Mobile-Cellular Networks for non-IP Related Activities, by Technology

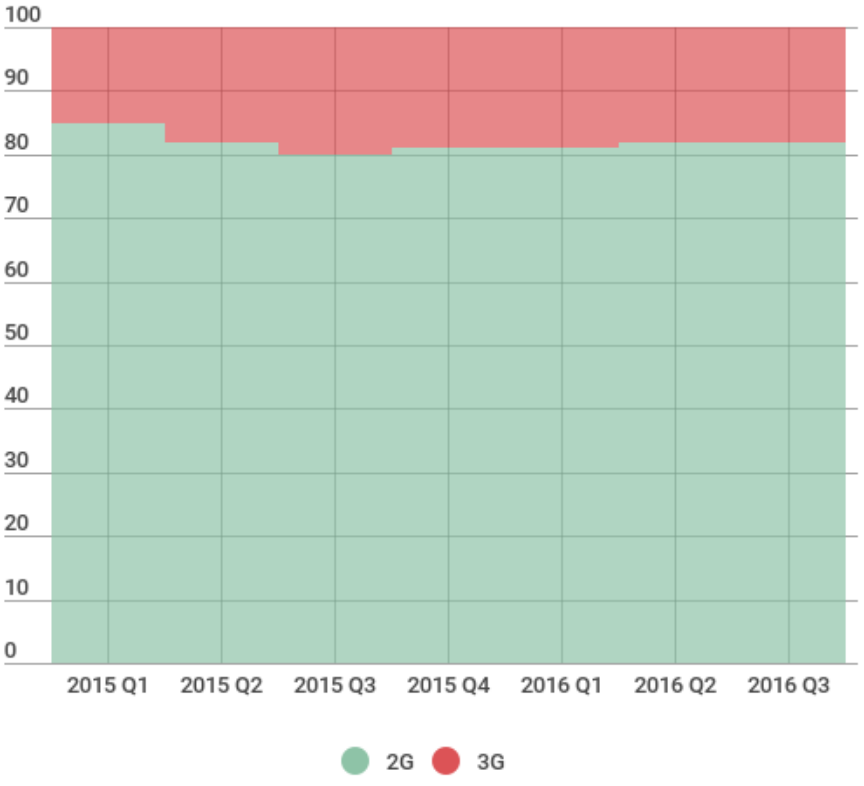


BD03: Urban Areas



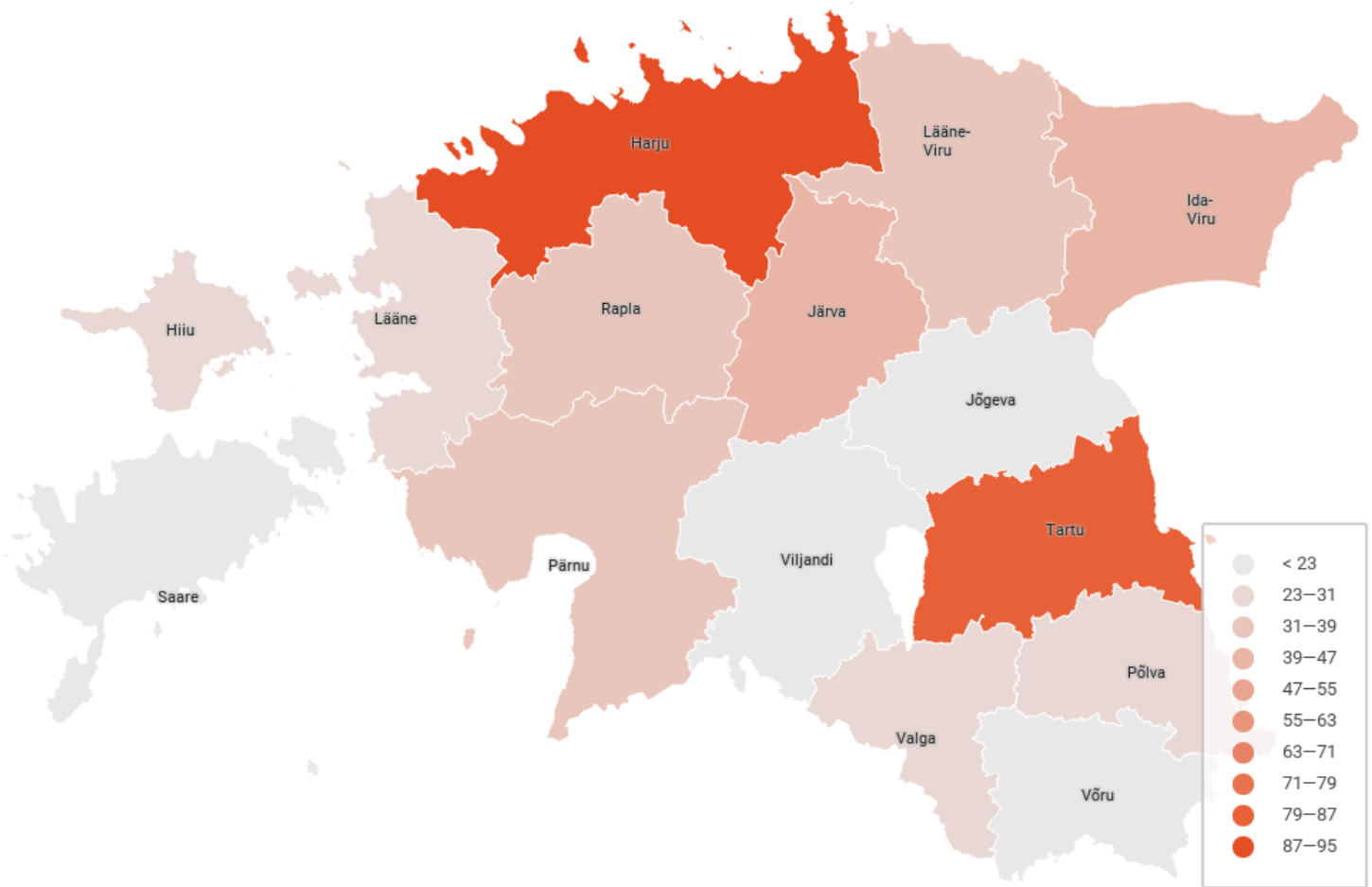
infogr.am

BD03: Rural Areas



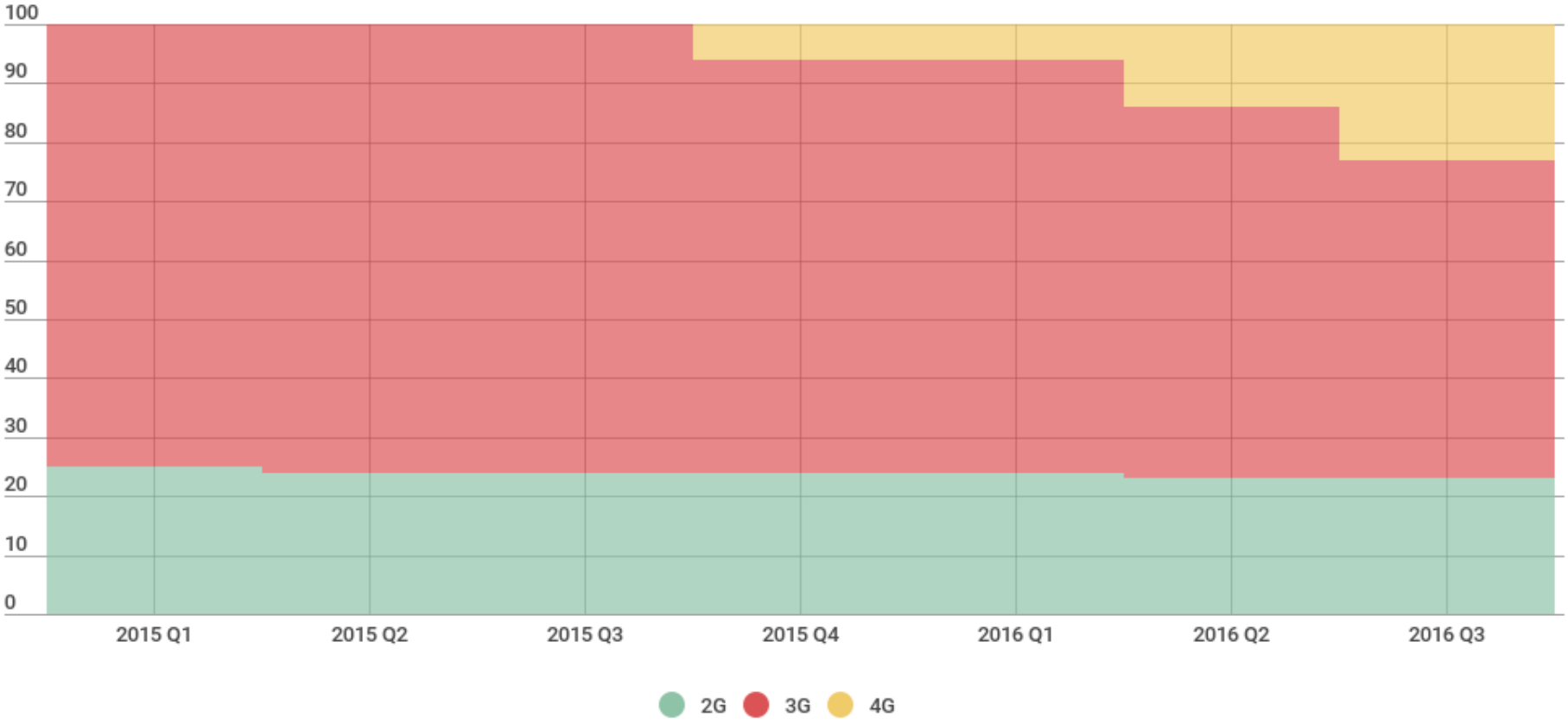
infogr.am

By Regions

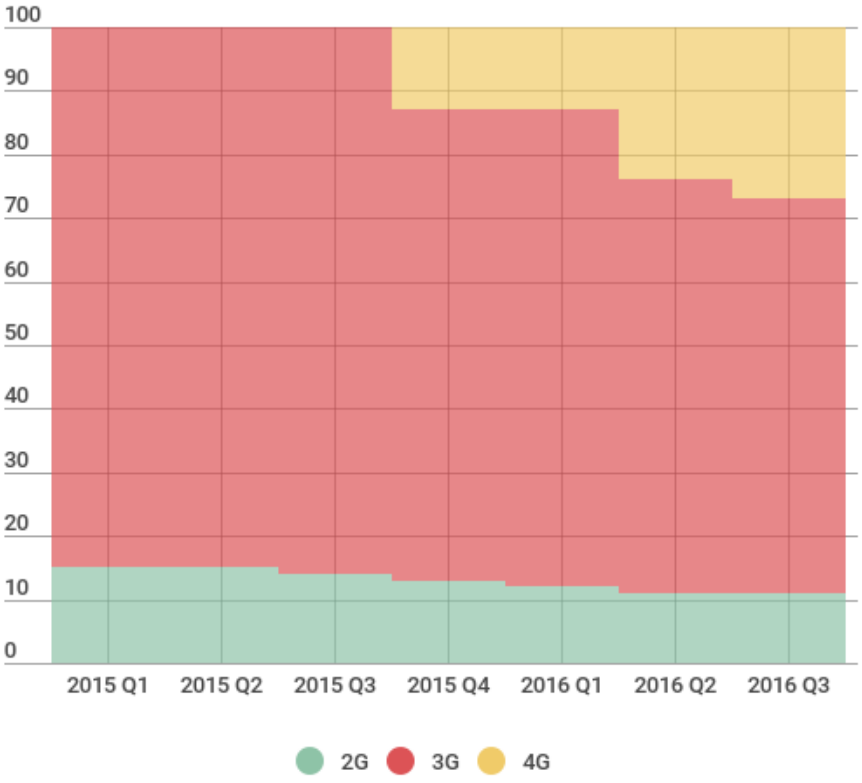


infogr.am

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

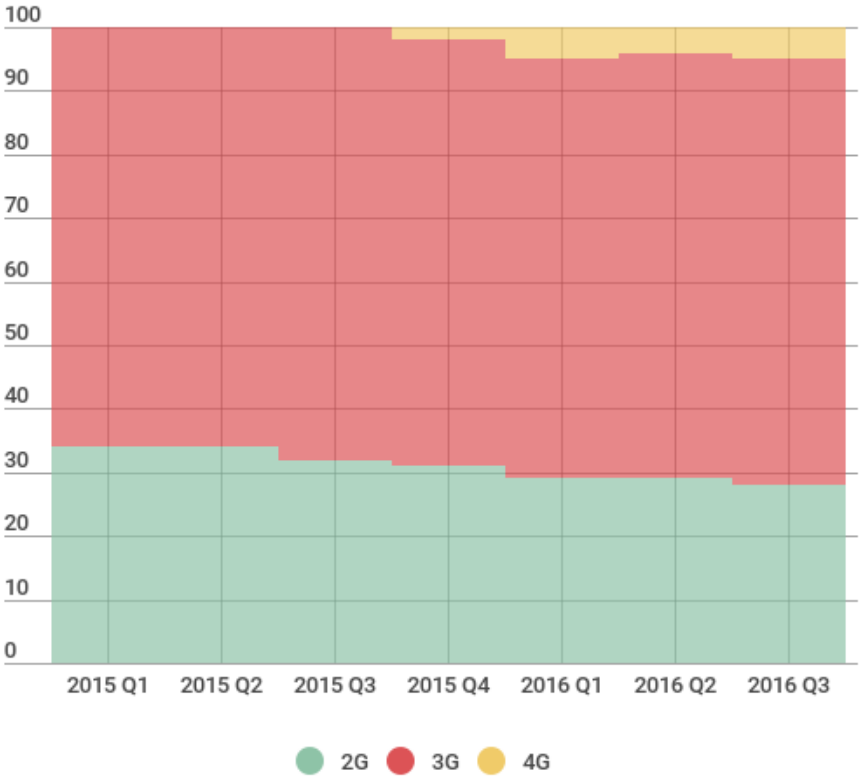


BD04: Urban Areas



infogr.am

BD04: Rural Areas



infogr.am

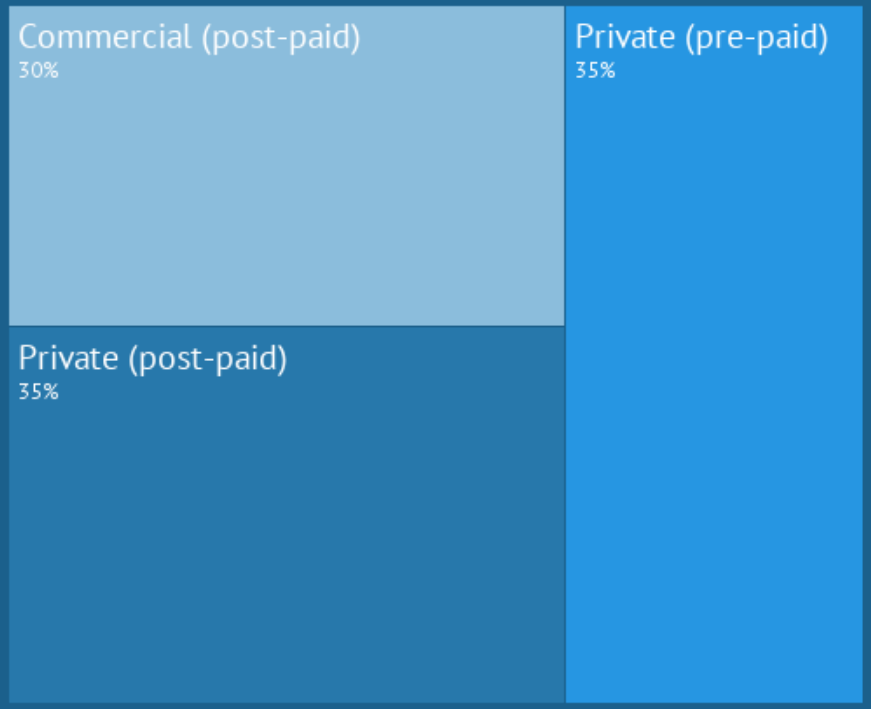
COMMERCIAL/PRIVATE CONTRACTS VS 4G USAGE (BD06-BD04)

Number of Subscribers (BD06)

infogr.am

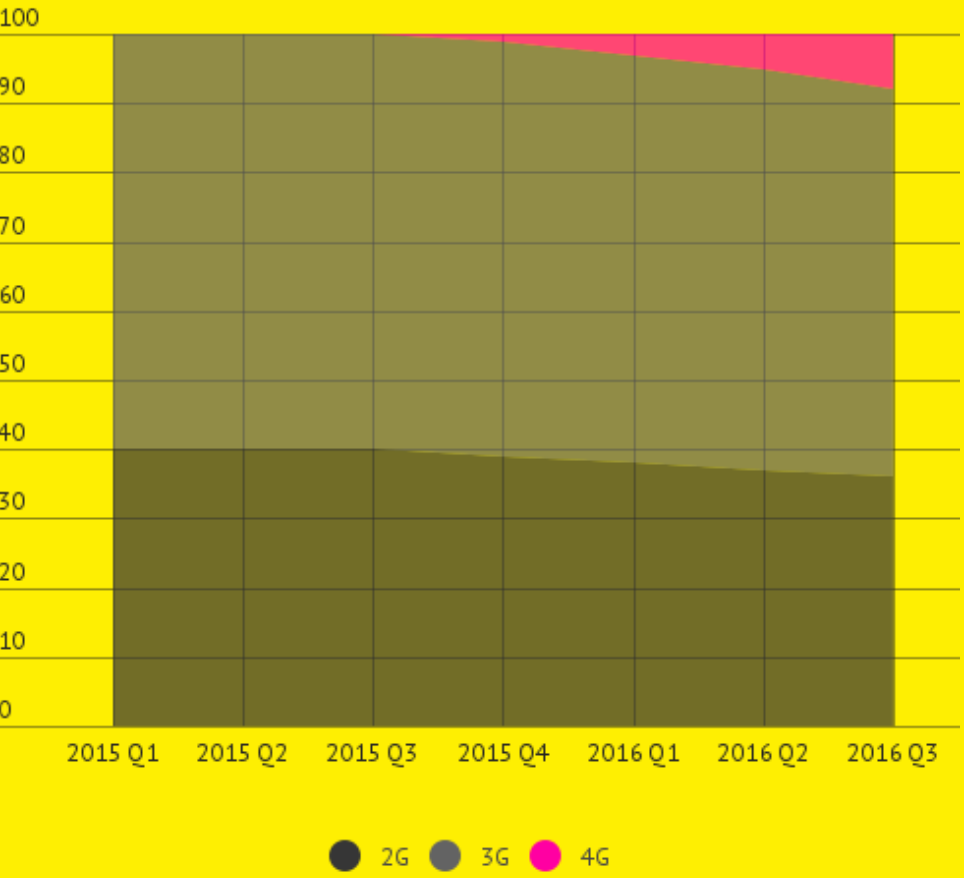
COMMERCIAL/PRIVATE CONTRACTS VS 4G USAGE (BD06-BD04)

Proportion of 4G Usage (BD04)



infogr.am

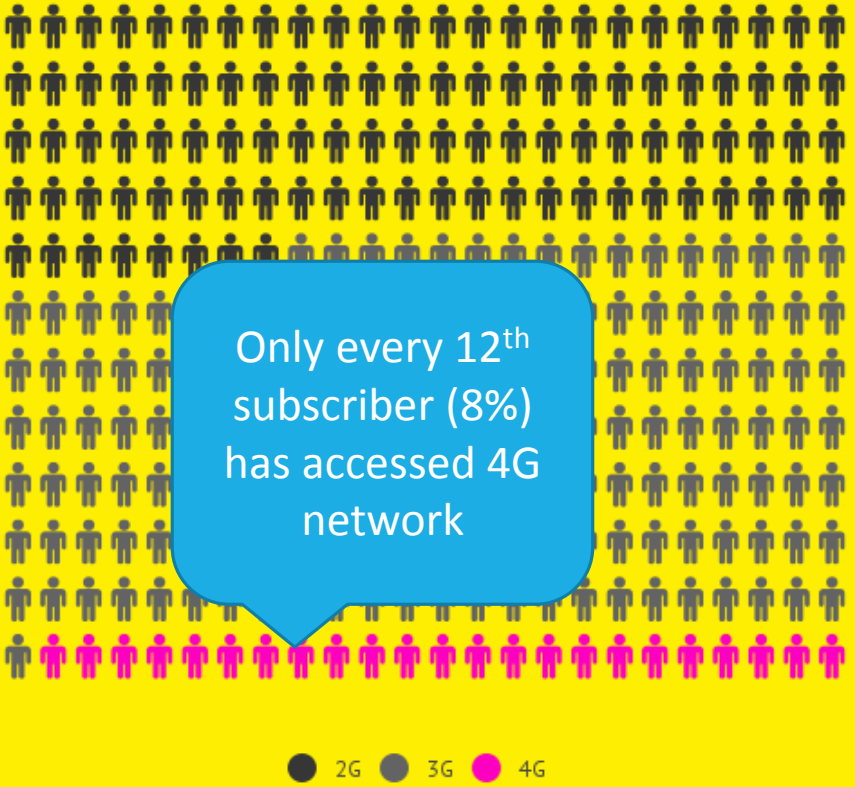
BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY



infogr.am

BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

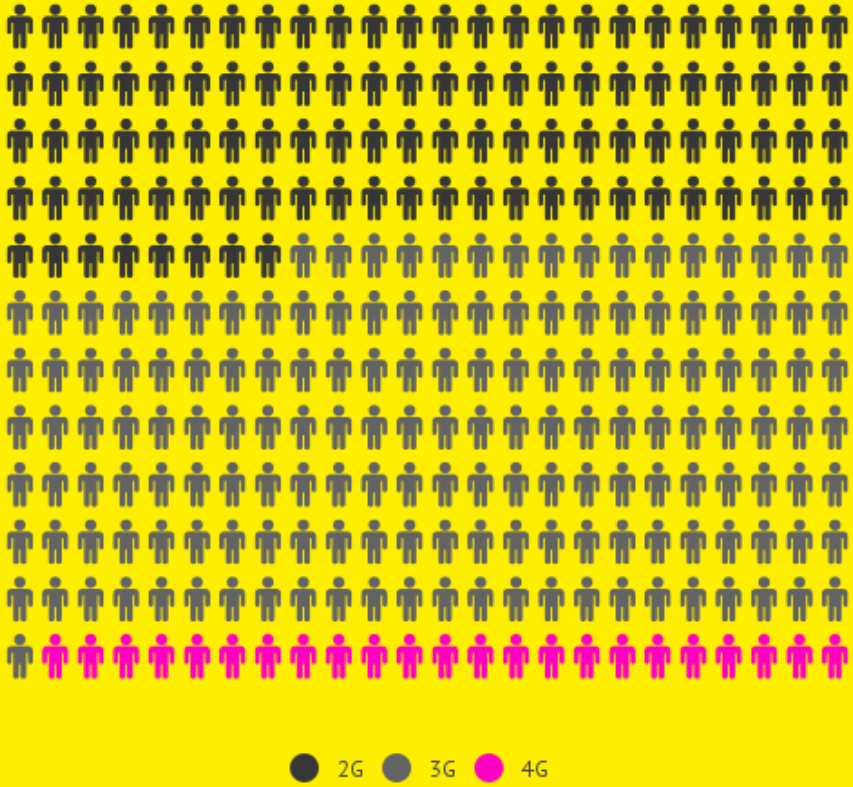
2016 Q3



infogr.am

BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

2016 Q3

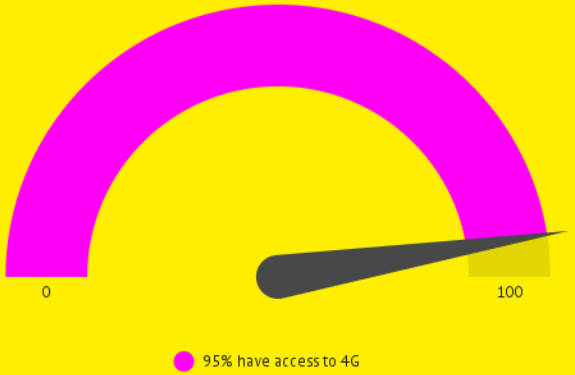


infogr.am

BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

Now

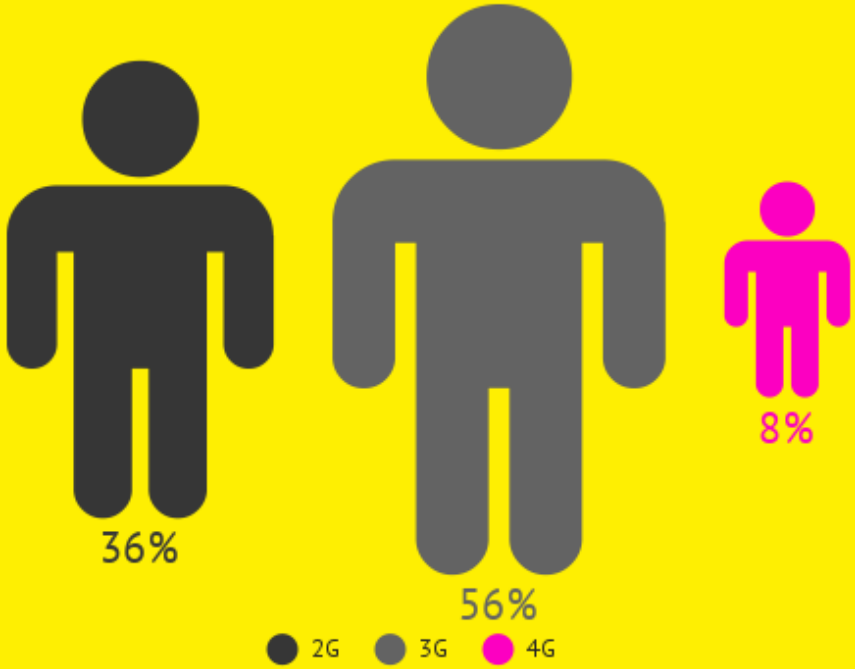
Objective



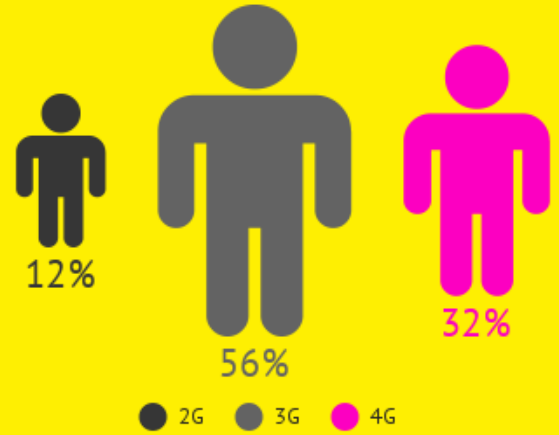
infogr.am

BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

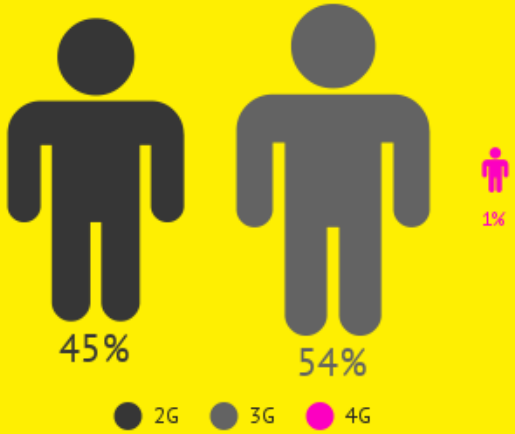
2016 Q3



Urban

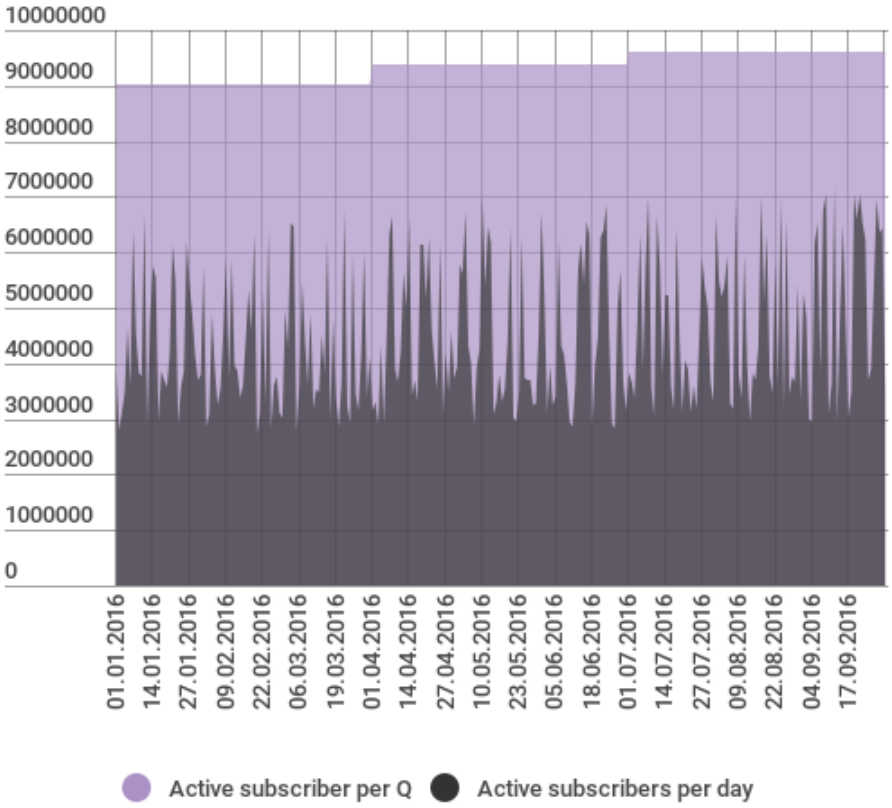


Rural



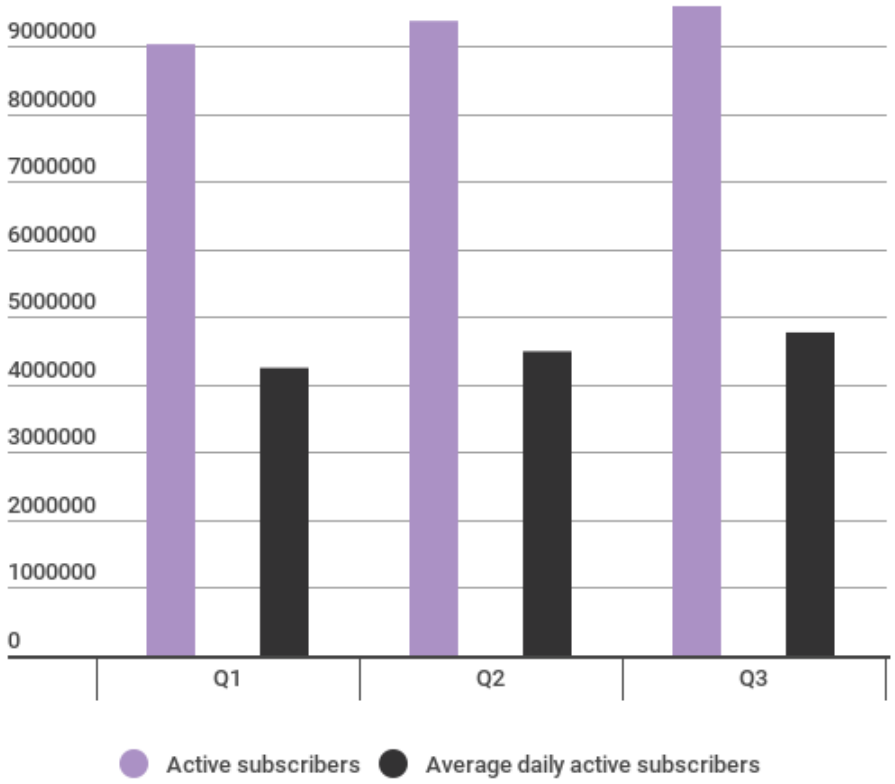
infogr.am

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



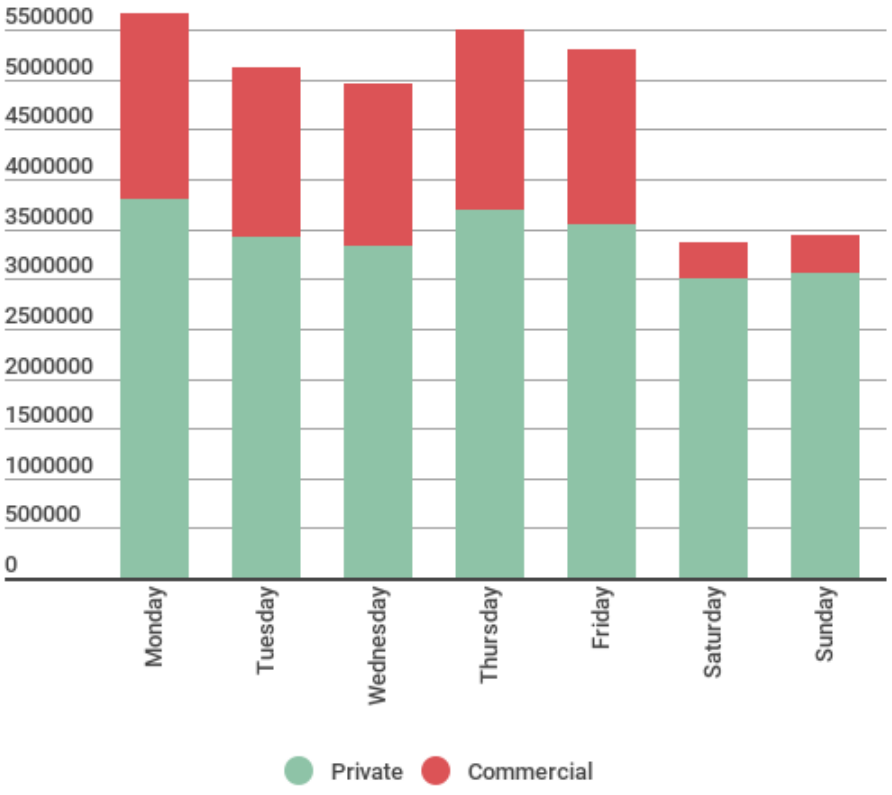
infogr.am

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



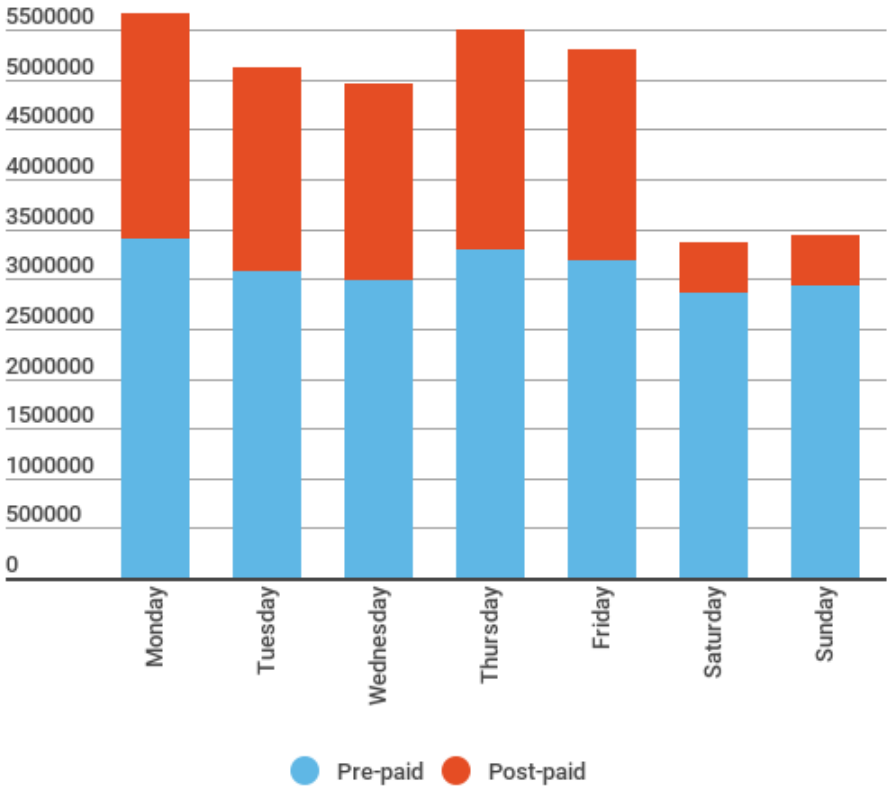
infogr.am

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



infogr.am

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



infogr.am

BD08: ACTIVE MOBILE DEVICES; BD09: IMEI CONVERSION RATE

Subscribers

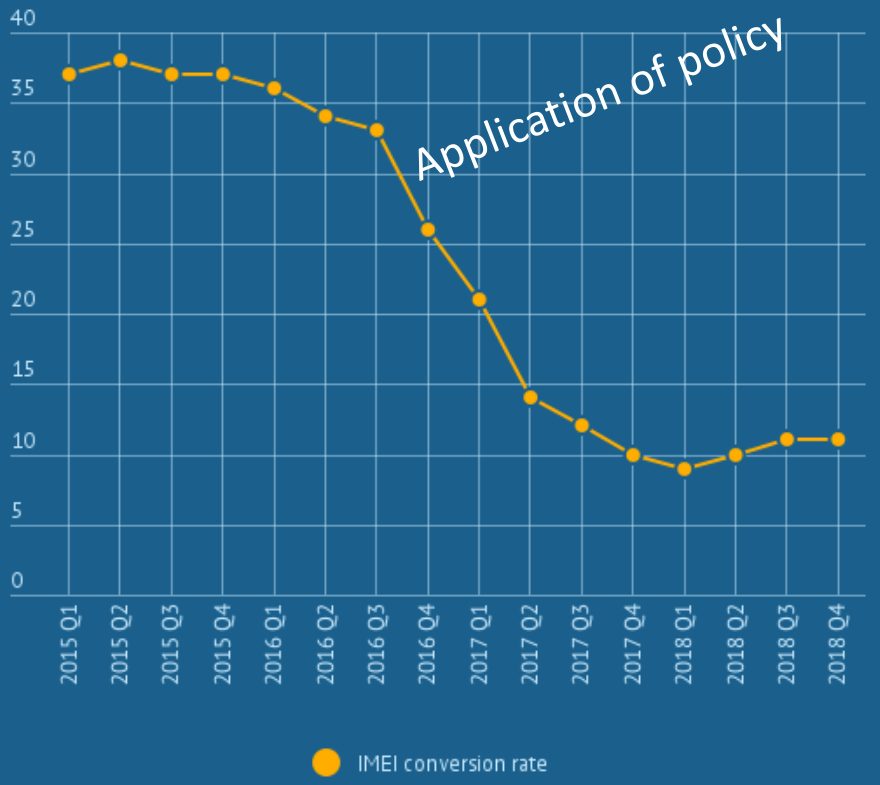
13,089,334

Number of active subscriptions 2016_Q3

25,262,406

Number of active devices 2016_Q3

1 device 2 or more devices



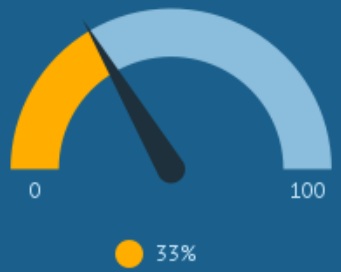
33%

Every third subscriber uses two or more devices with one SIM card. This suggests there is a "black market" for stolen devices.

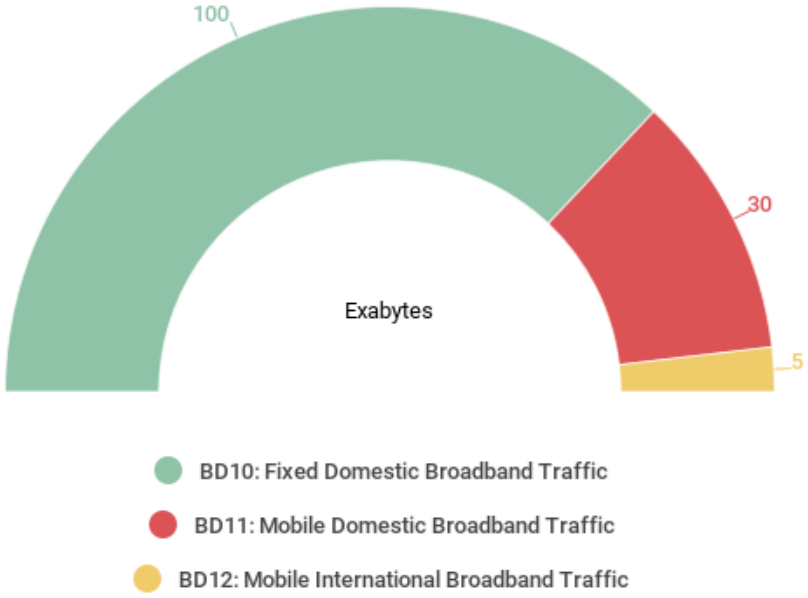
10%

Normal IMEI conversion rate would indicate how many subscribers obtain new phone per quarter.

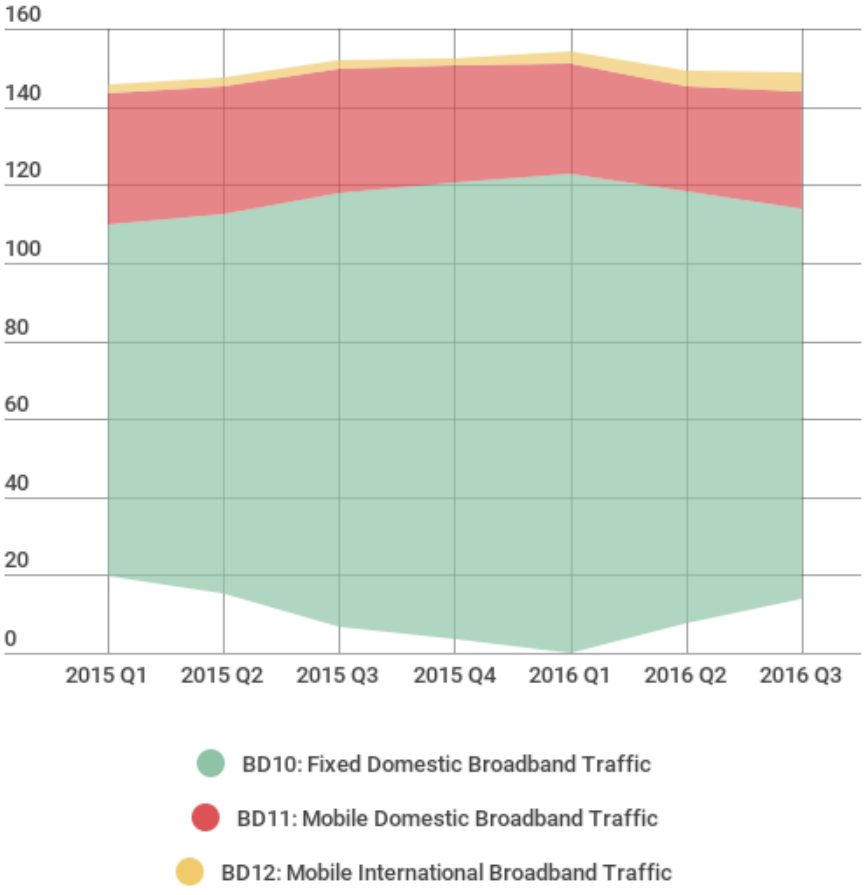
2016 Q3



2017 Q4

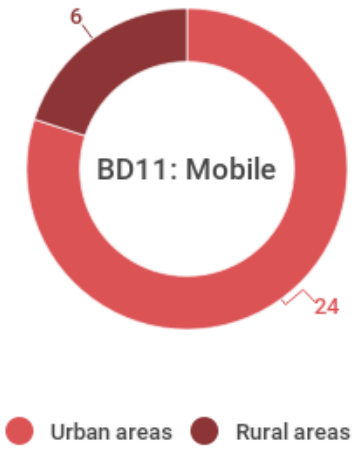


BD10, BD11, BD12: Broadband Traffic



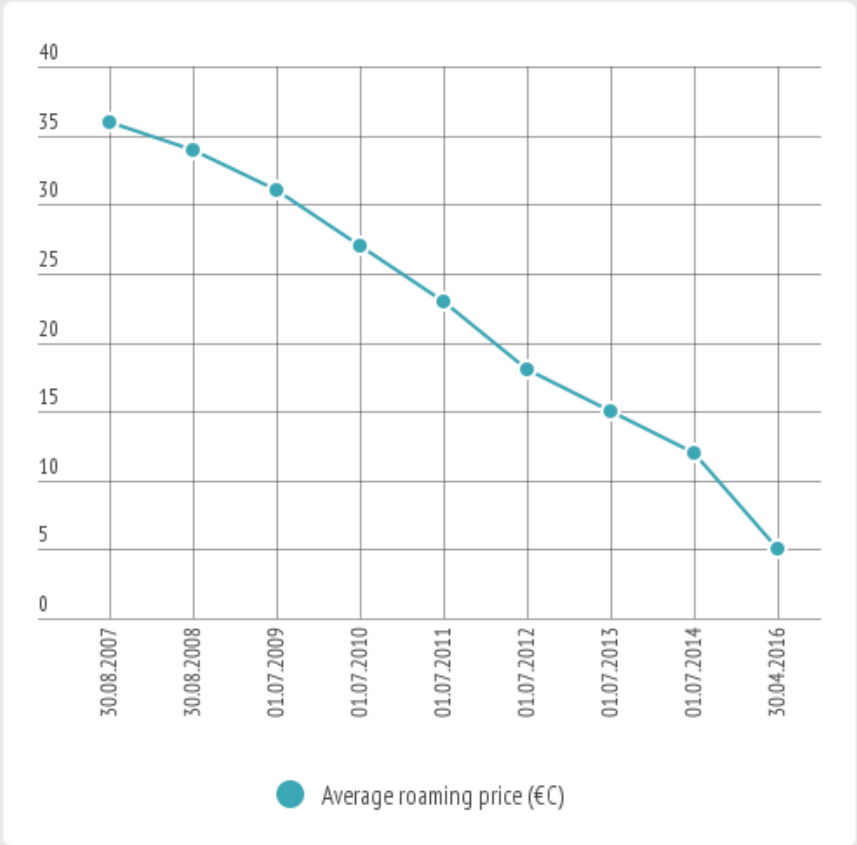
BD10: Fixed

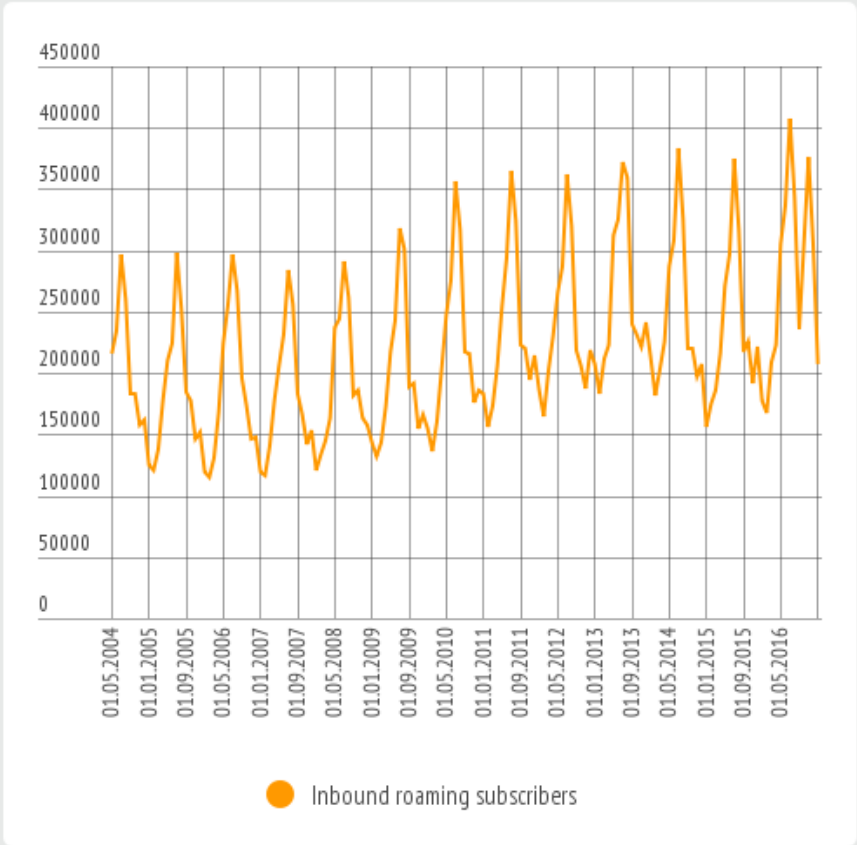
BD11: Mobile

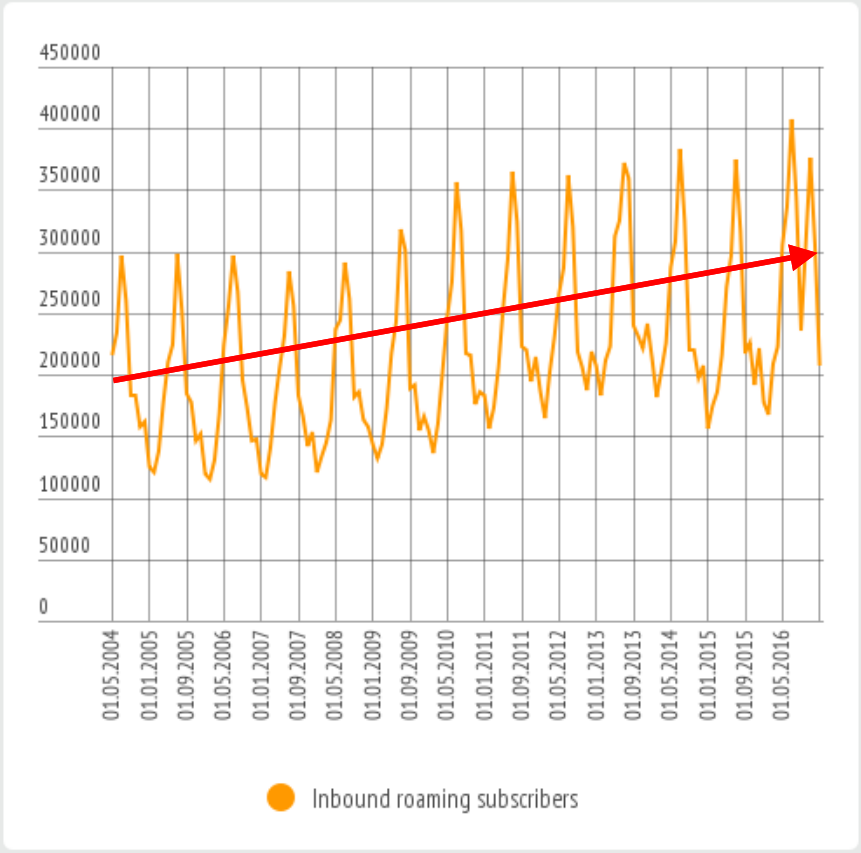


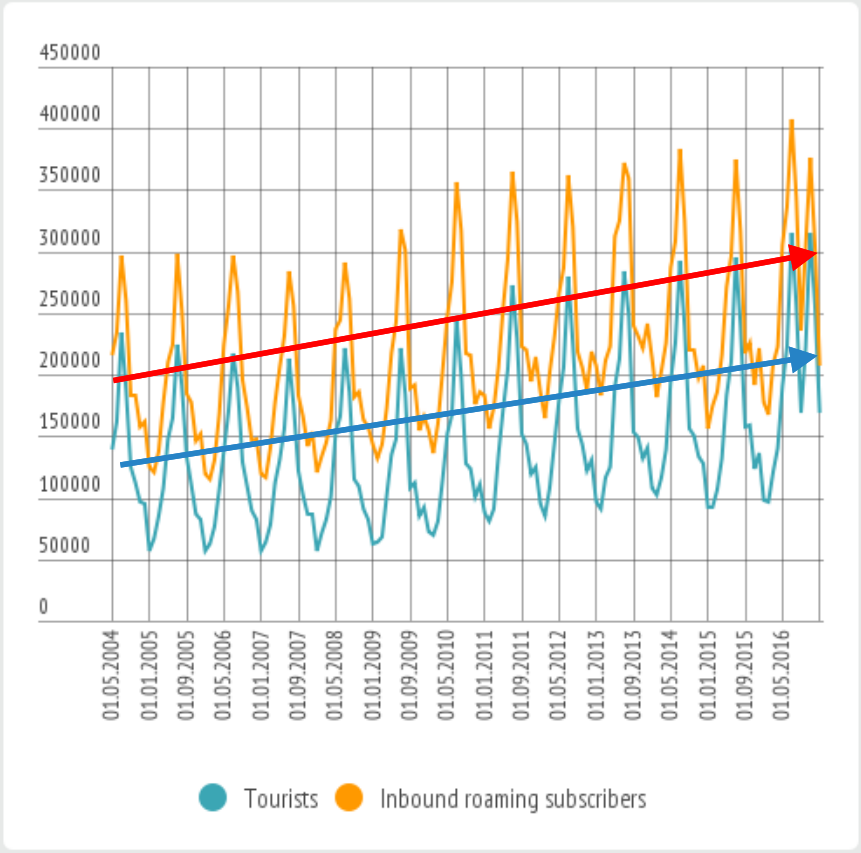
infogr.am

BD13: Inbound Roaming Subscriptions per Foreign Tourist. Effect of the Roaming Regulations



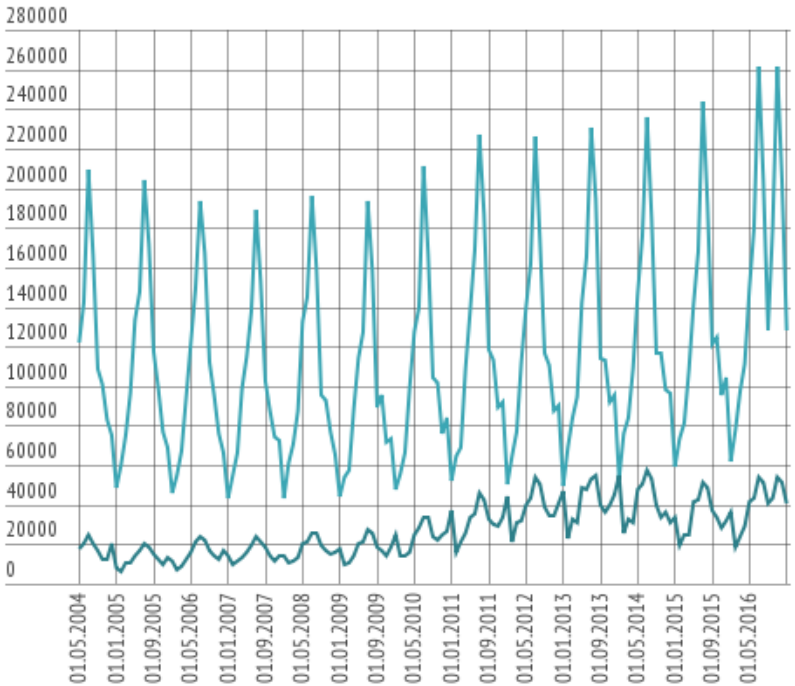


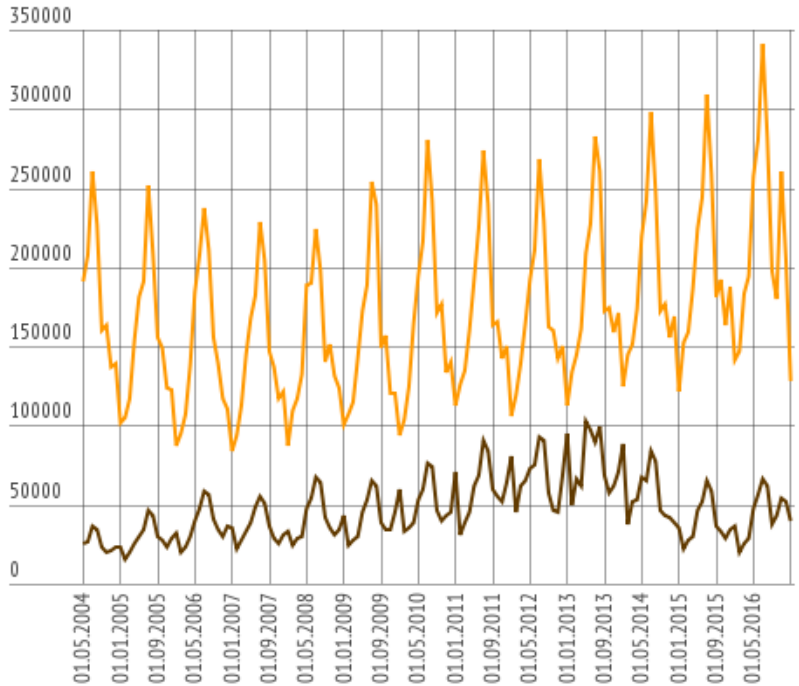




infogr.am

infogr.am

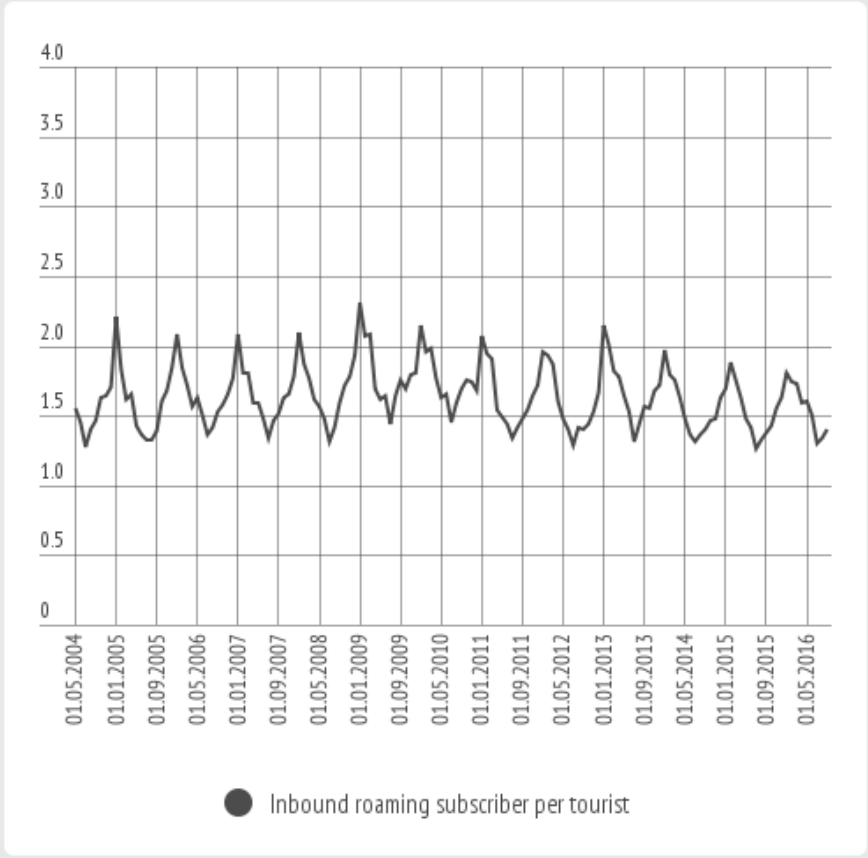




infogr.am

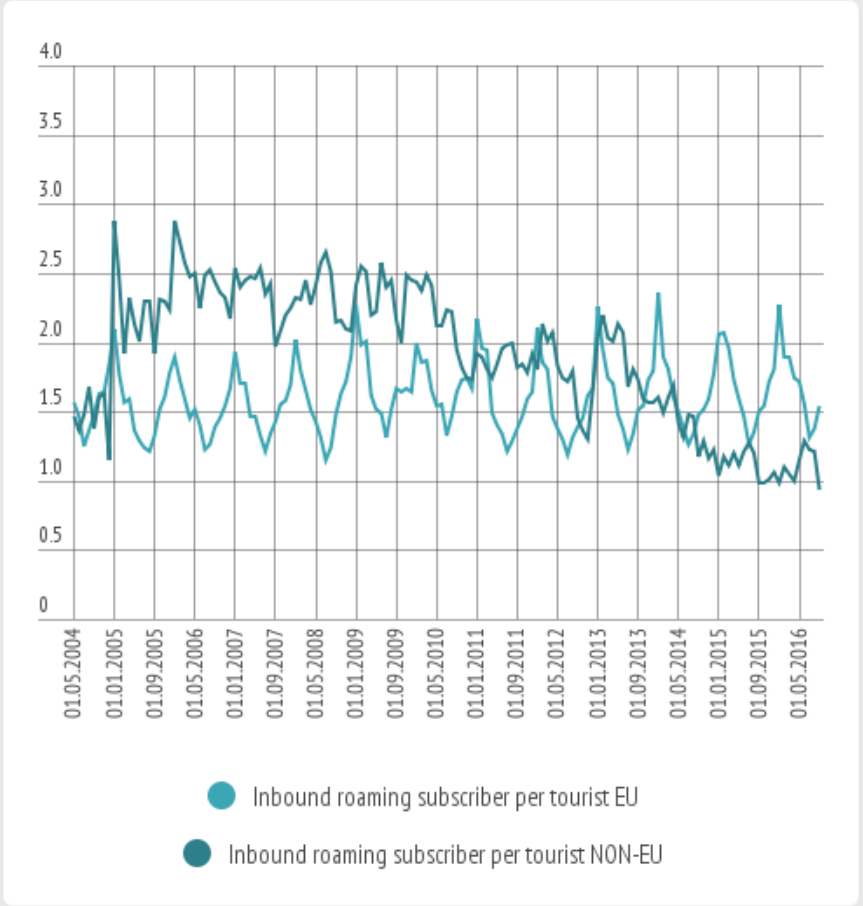
infogr.am

BD13: Inbound Roaming Subscriptions per Foreign Tourist



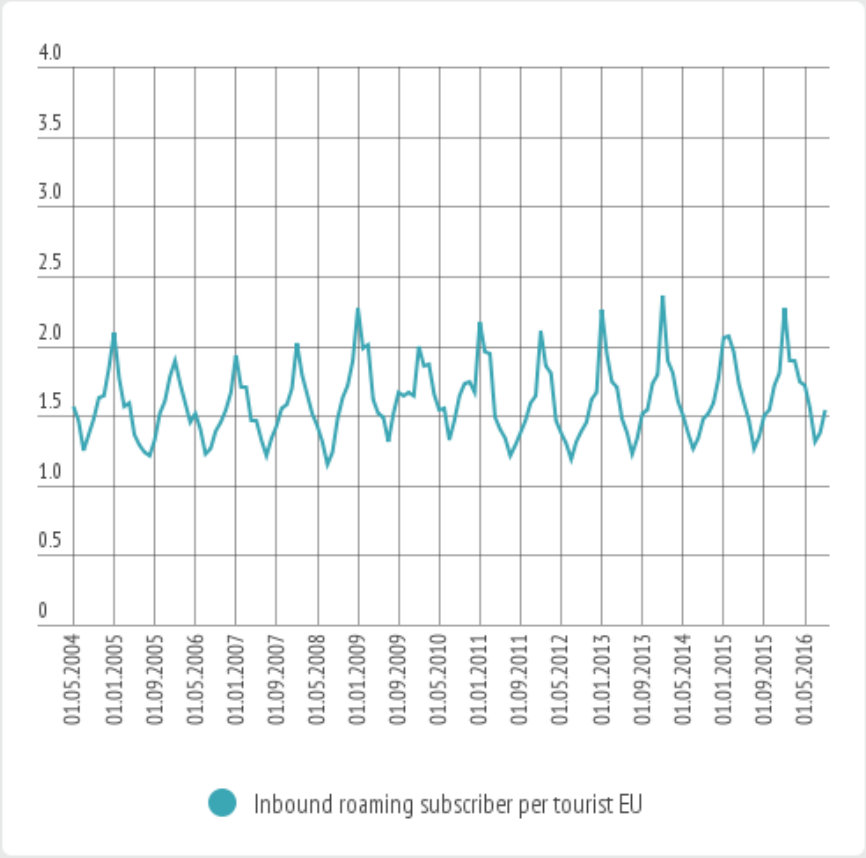
infogr.am

BD13: Inbound Roaming Subscriptions per Foreign Tourist



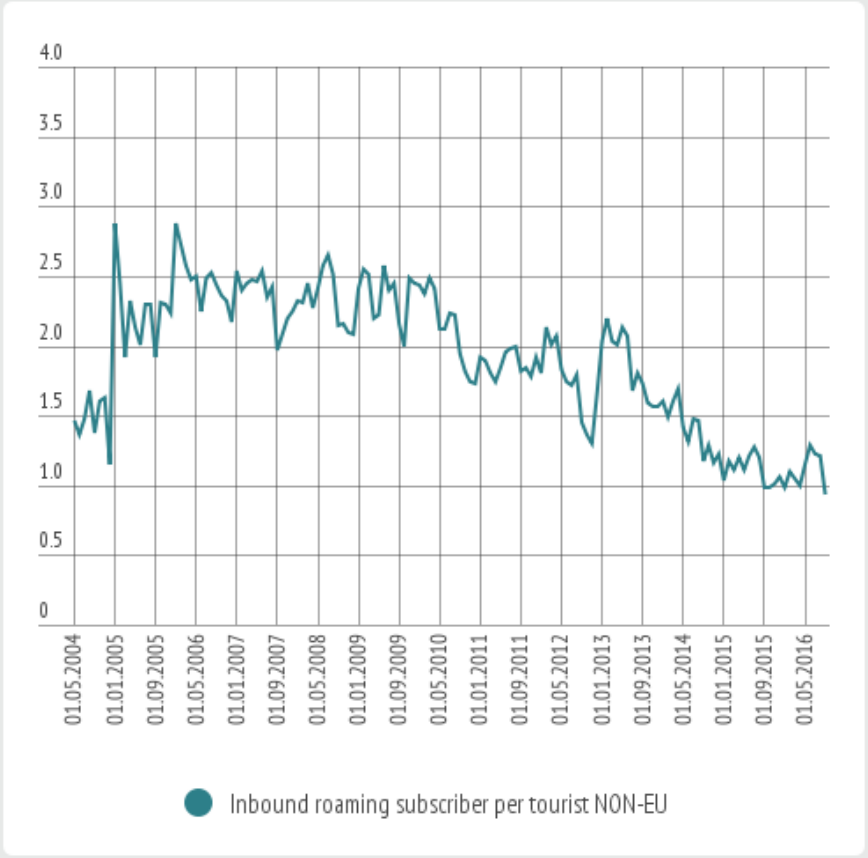
infogr.am

BD13: Inbound Roaming Subscriptions per Foreign Tourist



infogr.am

BD13: Inbound Roaming Subscriptions per Foreign Tourist



infogr.am

BD16+

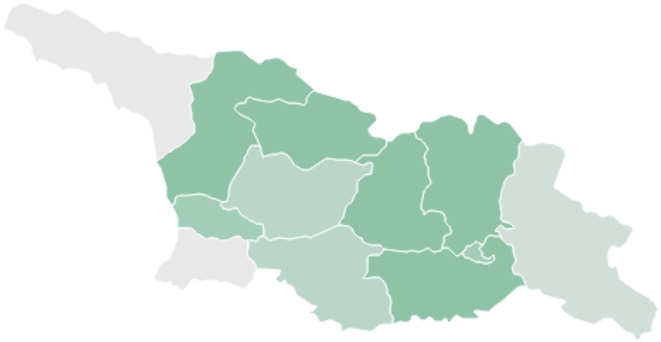
Pilot countries have an option to include additional indicators that are relevant to the country:

- Kenya: Including indicators on Mobile Money
- UAE: Interested in some data mining and machine learning approaches
- Colombia: Density of population census block level
- Stakeholders in other countries welcome to include indicators that they feel are necessary from the point of view of their country and globally

Internationally comparable indicators

Difference with Current Reporting System

- Breakdown
  - Geographical: Local Administrative Units 1-2-3
  - Urban/Rural breakdown
  - Contract type: private/commercial; pre-paid/post-paid; voice/data
  - Socio-demographics: gender, age, language
  - Mobile technology generation: 2G, 3G, 4G
  - Fixed technology: cable, DSL, fibre, etc.
  - Fixed advertised speed
  - Device based on IMEI/TAC
  - Event type (call, message, incoming, outgoing, IP)
  - Data volume



- Data providers don’t simply report summarised indicators, but need to calculate and aggregate using Big Data tools

Comments from the Country Stakeholders

Excited that this project has started and they are looking for the results and want to commit to the project

Asking if the project will also have influence on the timeliness of the reporting and publication of the ICT indicators locally and globally (via ITU)

Although the pilot project is a starting point and therefore limited by its scope, stakeholders expect more big data related activities initiated locally and on global scale in terms of tools, possibilities of big data, capacity building, and development of the resources:

- Data mining, data discovery
- Machine learning
- Improvements of ICT, SDG, eliminating digital divide, poverty
- Capacity building, academic research

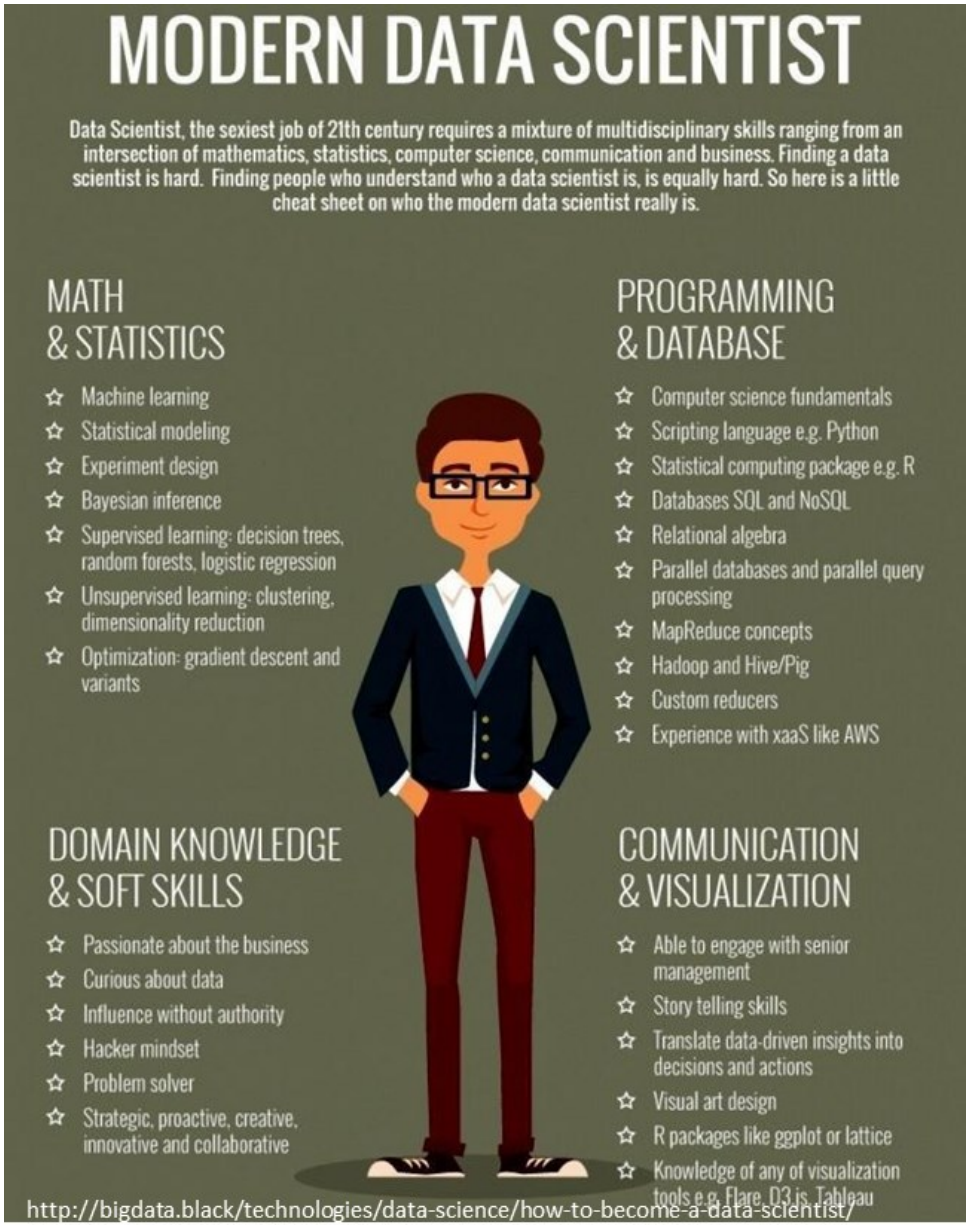
Data Scientist

To assist pilot countries to process the data

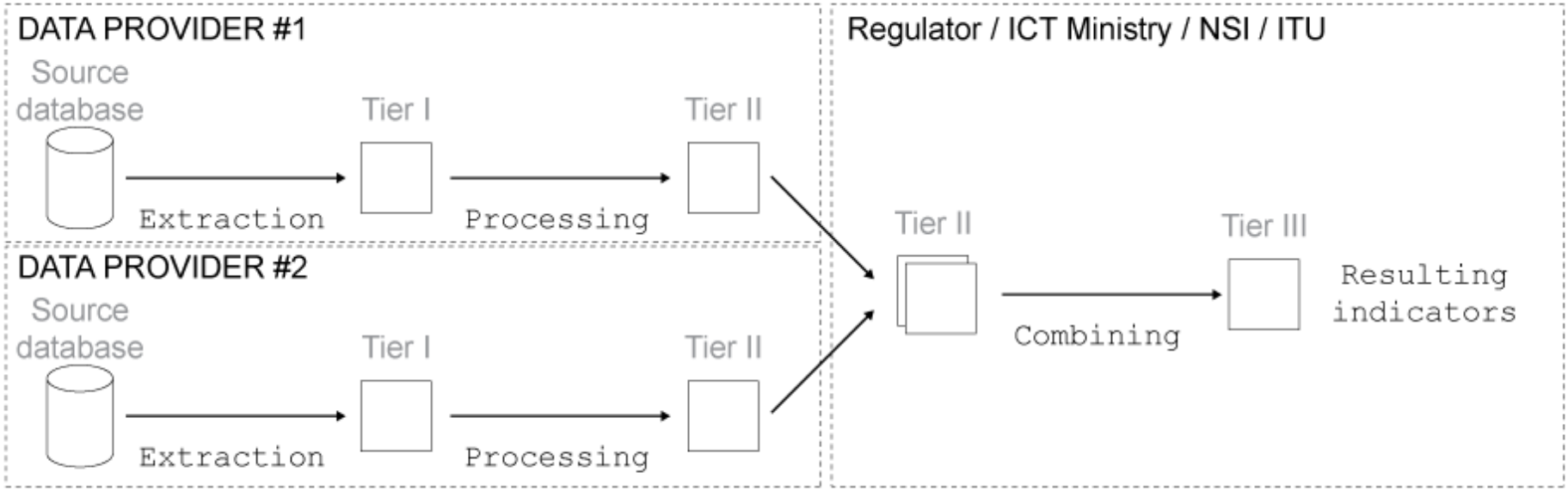
Combining the data providers' indicators into country indicators

Analyse the indicators for the countries

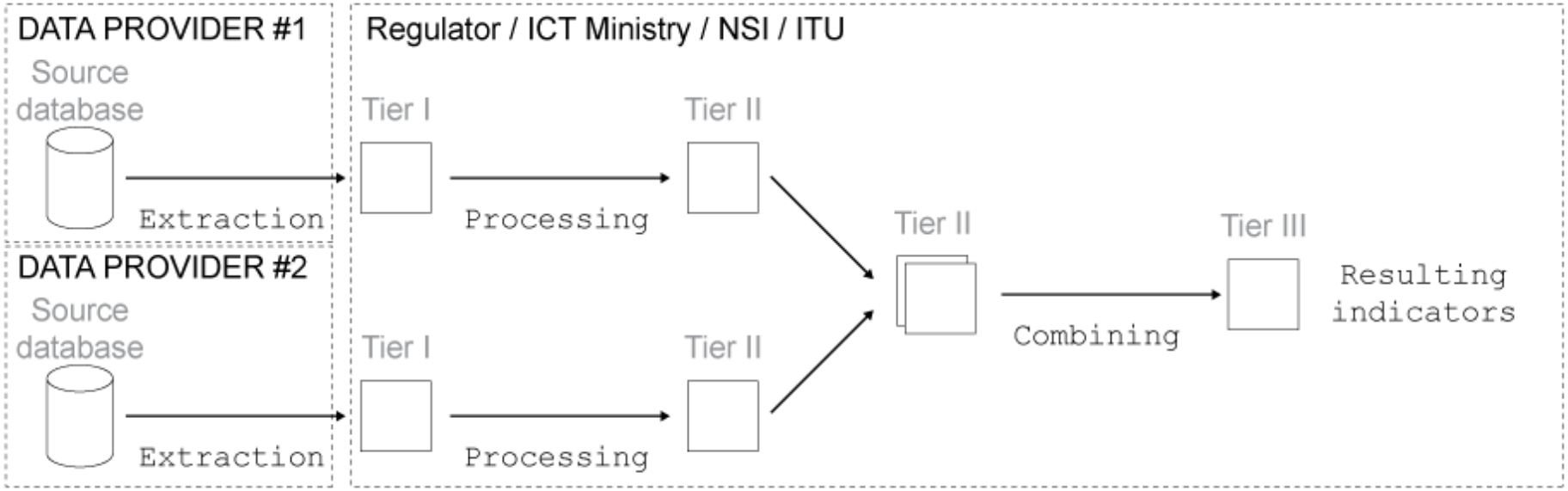
Visualisation and country reports



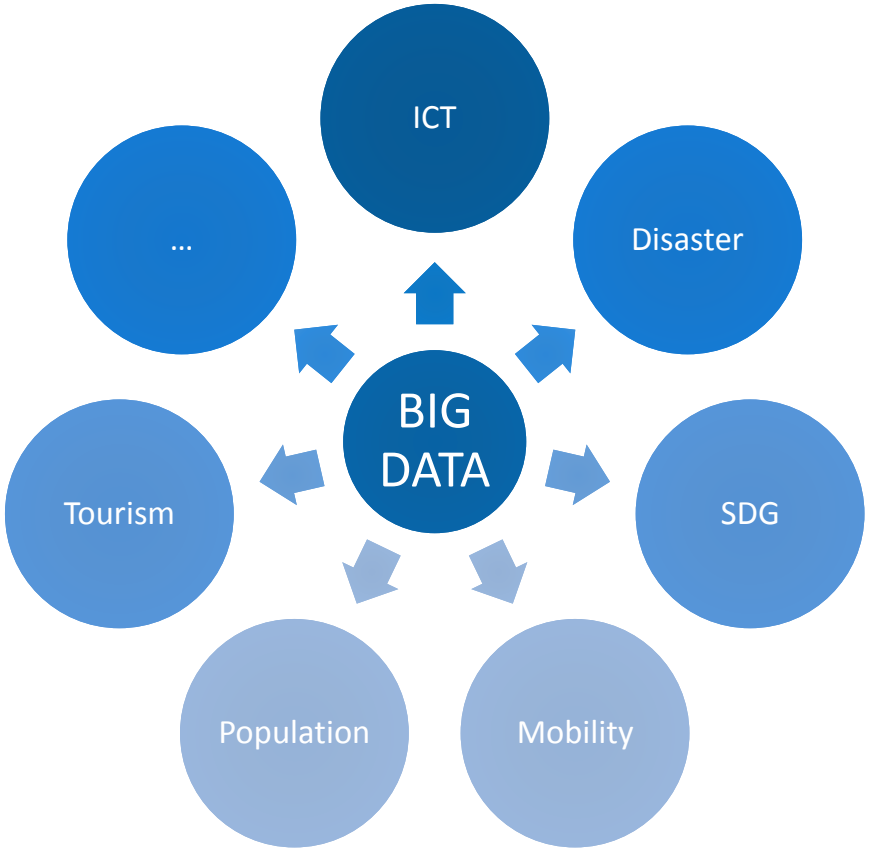
Processing Location: Option 1

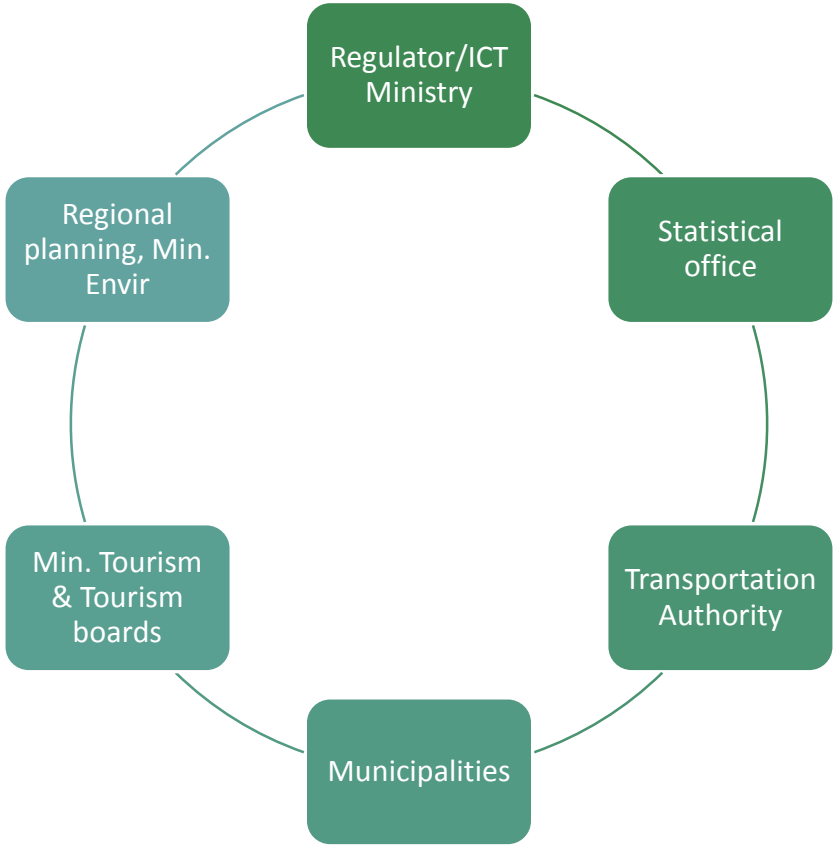


Processing Location: Option 2



Possibility of common use of the big data sources for





Project Results

The expected results of the project are:

- 6 country reports
- Concrete ICT indicators database on each pilot country
- Description of the source data and methodology for processing of the data
- Possibilities and limitations of the data source and indicators
- How the indicators can be used for policy and investment decisions
- Project final report

Based on the pilot studies, further discussion about using such indicators globally will continue, Phase II

Presentation of project:

- EGH, EGTI meetings October 2016
- WTIS 2016 - World Telecommunication/ICT Indicators Symposium November 2016
- Next conferences on ICT and big data

Margus Tiru

Consultant to ITU

project coordinator

margus.tiru@gmail.com

Esperanza Magpantay

Senior Statistician, ITU ICT Data and Statistics Division

esperanza.magpantay@itu.int

Susan Teltscher

Head of the ICT Data and Statistics Division

ITU Telecommunication Development Bureau

susan.teltscher@itu.int