



ITU Regional Workshop on ICT Statistics

Almaty, Republic of Kazakhstan

31 January - 1 February 2018



Big data for measuring the information society

Ivan Vallejo Vall

ICT Data and Statistics Division

International Telecommunication Union

Index

- **Relevance of big data from the ICT sector**
- Big data for measuring the information society
- Experience in CIS countries – tour de table



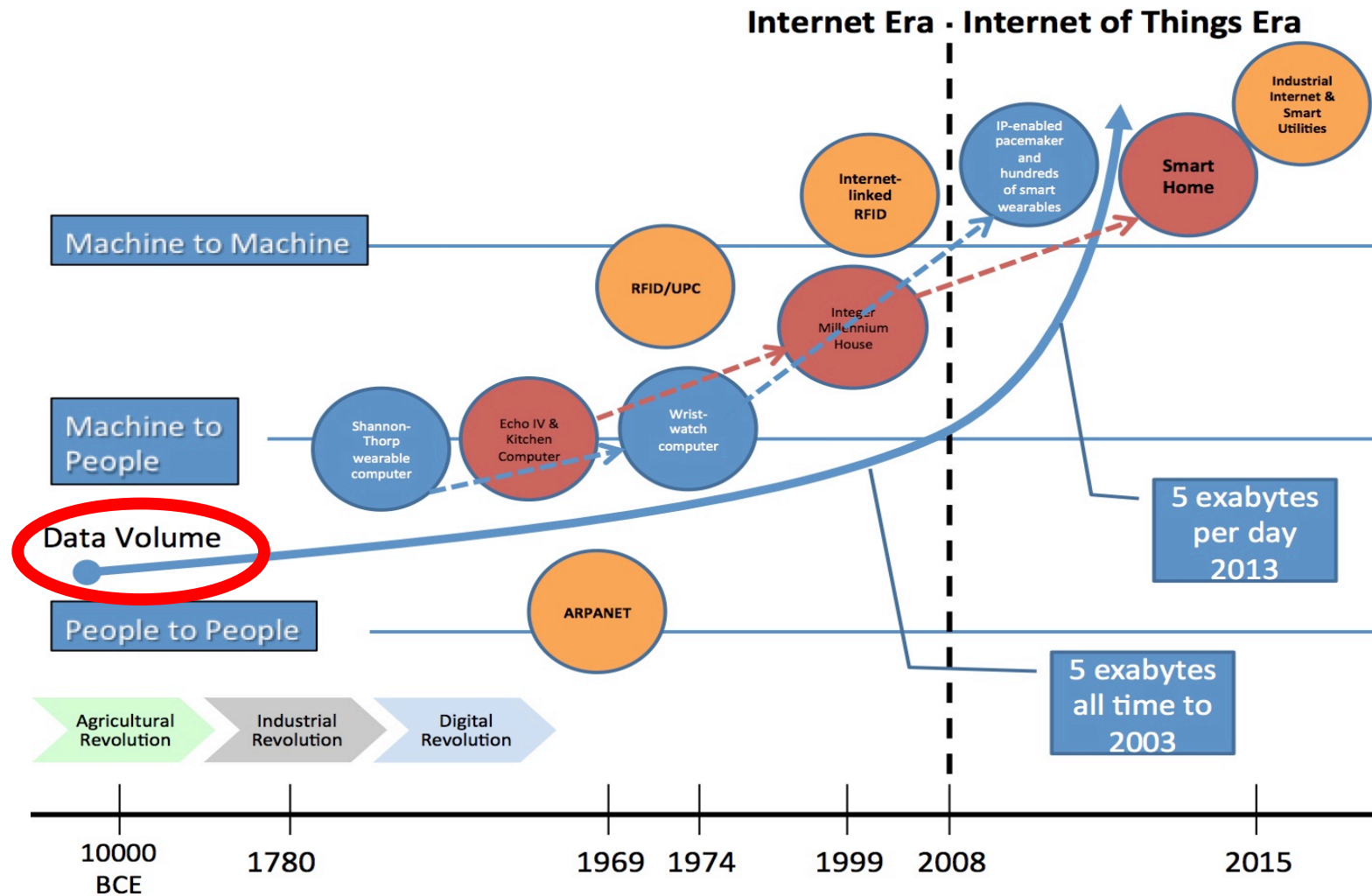
Big data in CIS countries

In your country:

1. Have you carried out any project to explore the use of big data from the ICT sector?
2. Are there plans to undertake projects in this area in the future?
3. Which are the main obstacles to using big data from the ICT sector for development and/or official statistics?

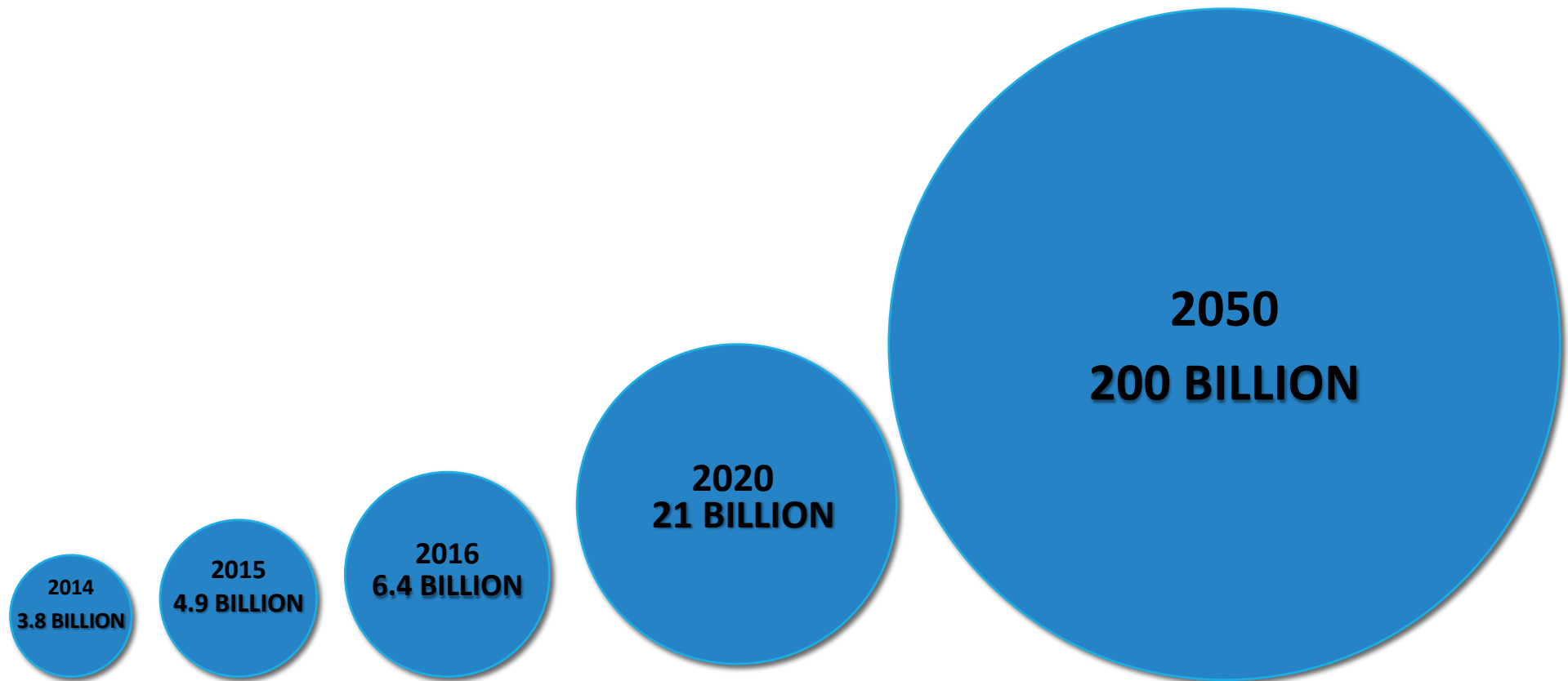


Big data growth

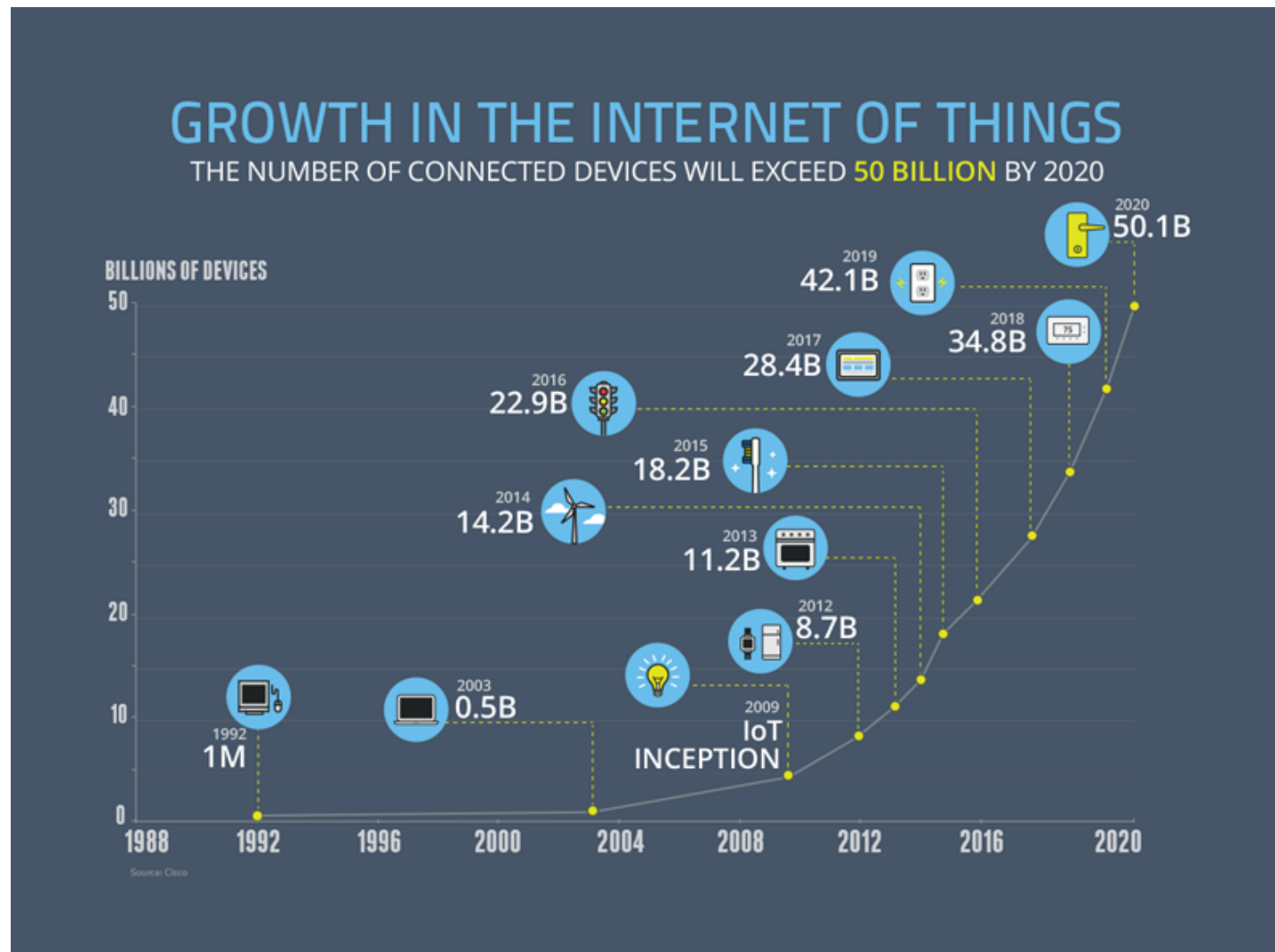




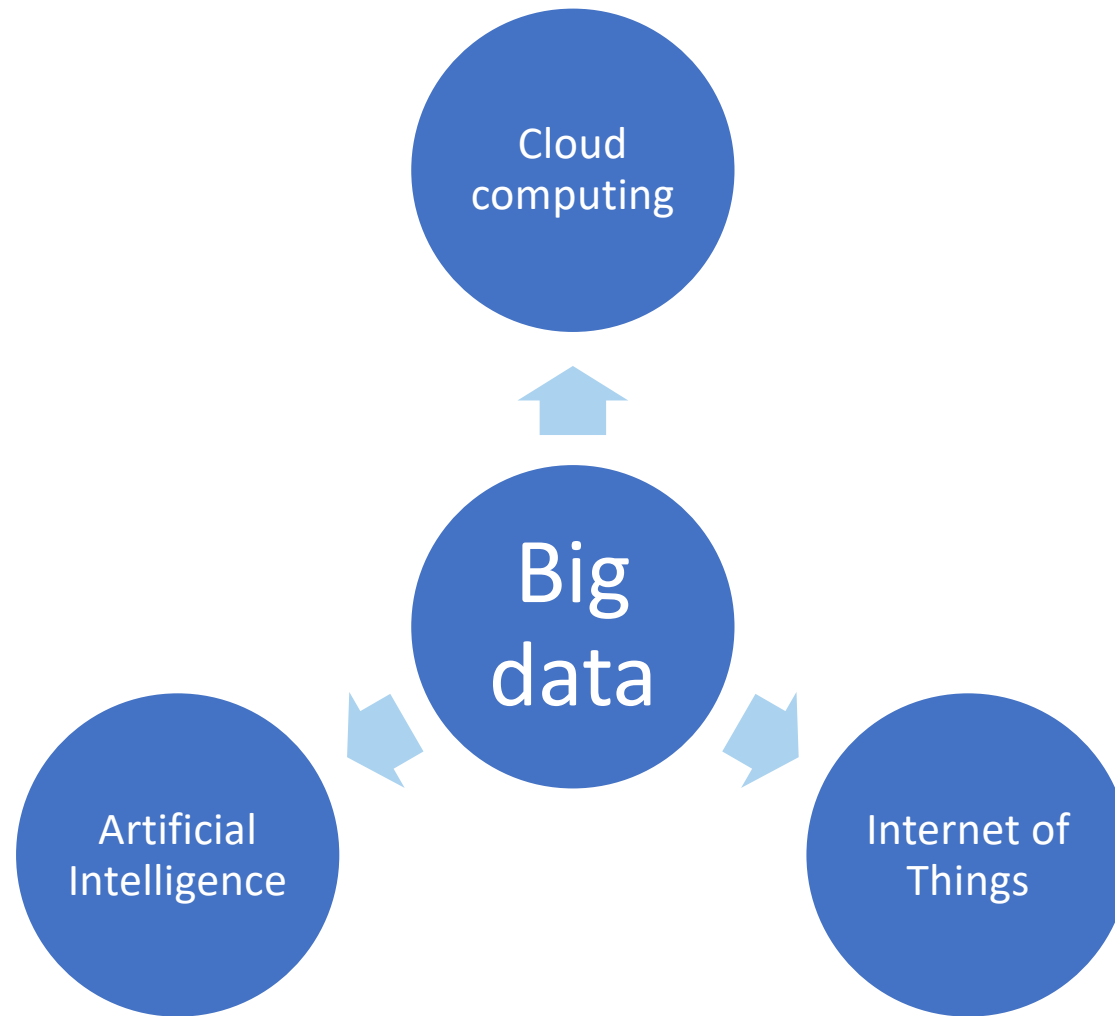
Number of connected devices



The Internet of Things



Emerging enabling technologies





ICT Sector – big data sources

Telecommunication
service providers

Fixed operators
Mobile operators
Internet service providers (ISPs)
Satellite companies

Internet and mobile
content providers

Over-the-top service providers (OTTs)
Social network providers
Mobile apps market/providers

Others

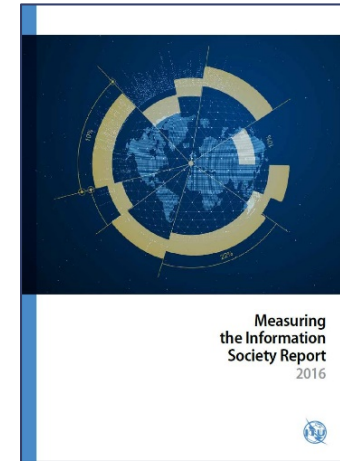
Software providers
Content distribution network (CDN)
providers
Equipment providers

ITU's engagement

Big data for official statistics



- UN agency for ICTs
- Measuring the Information Society Reports – big data and IoT data
- WTIS – panel debates
- Discussion items in EGH and EGTI
- ITU projects: preventing the spread of epidemics (Guinea, Liberia, Sierra Leone)
- Member of UN Global Working Group (GWG) on Big Data for Official Statistics
- **Project on Big Data for Measuring the Information Society**



Index

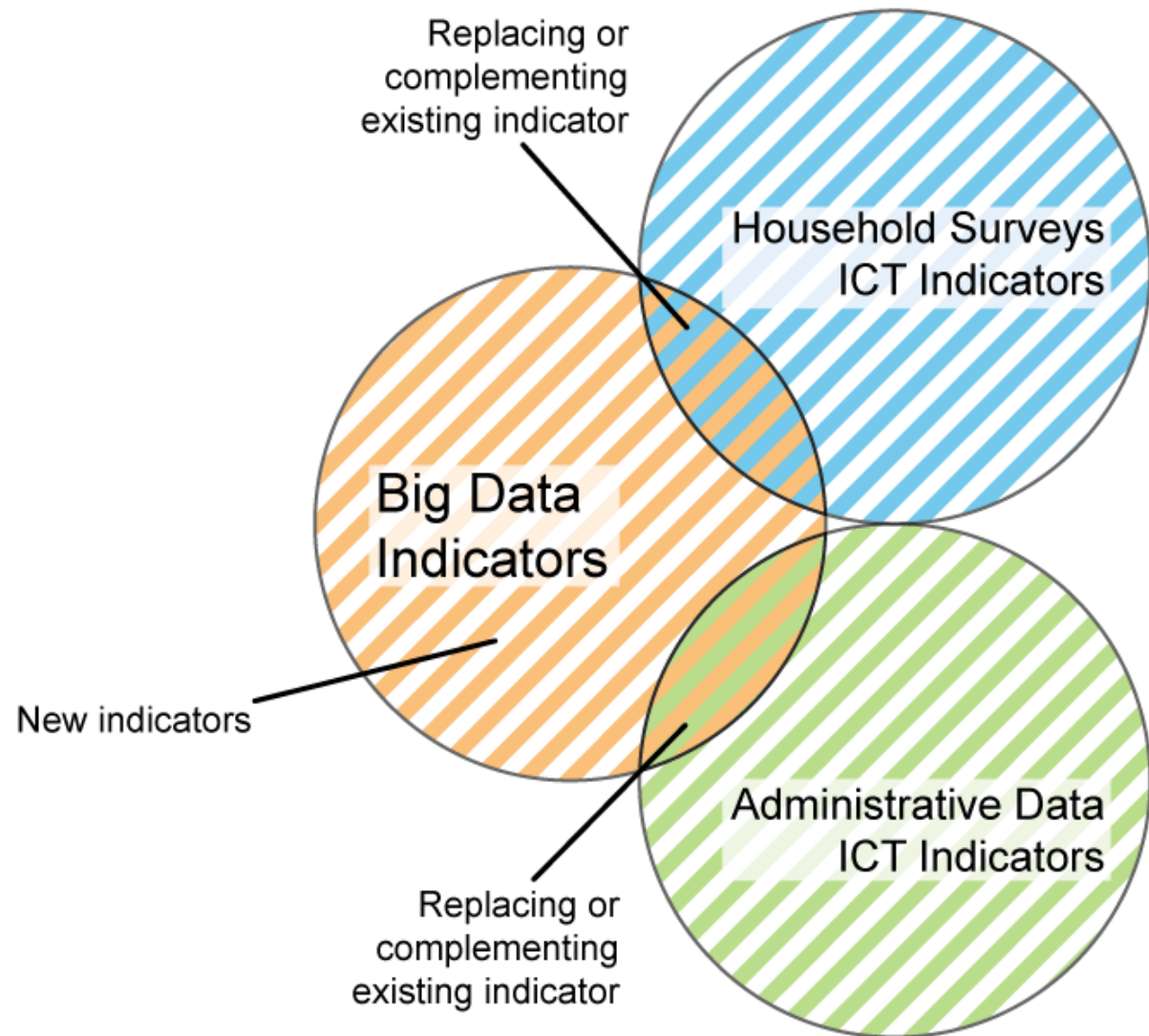
- Relevance of big data from the ICT sector
- **Big data for measuring the information society**
- Experience in CIS countries – tour de table

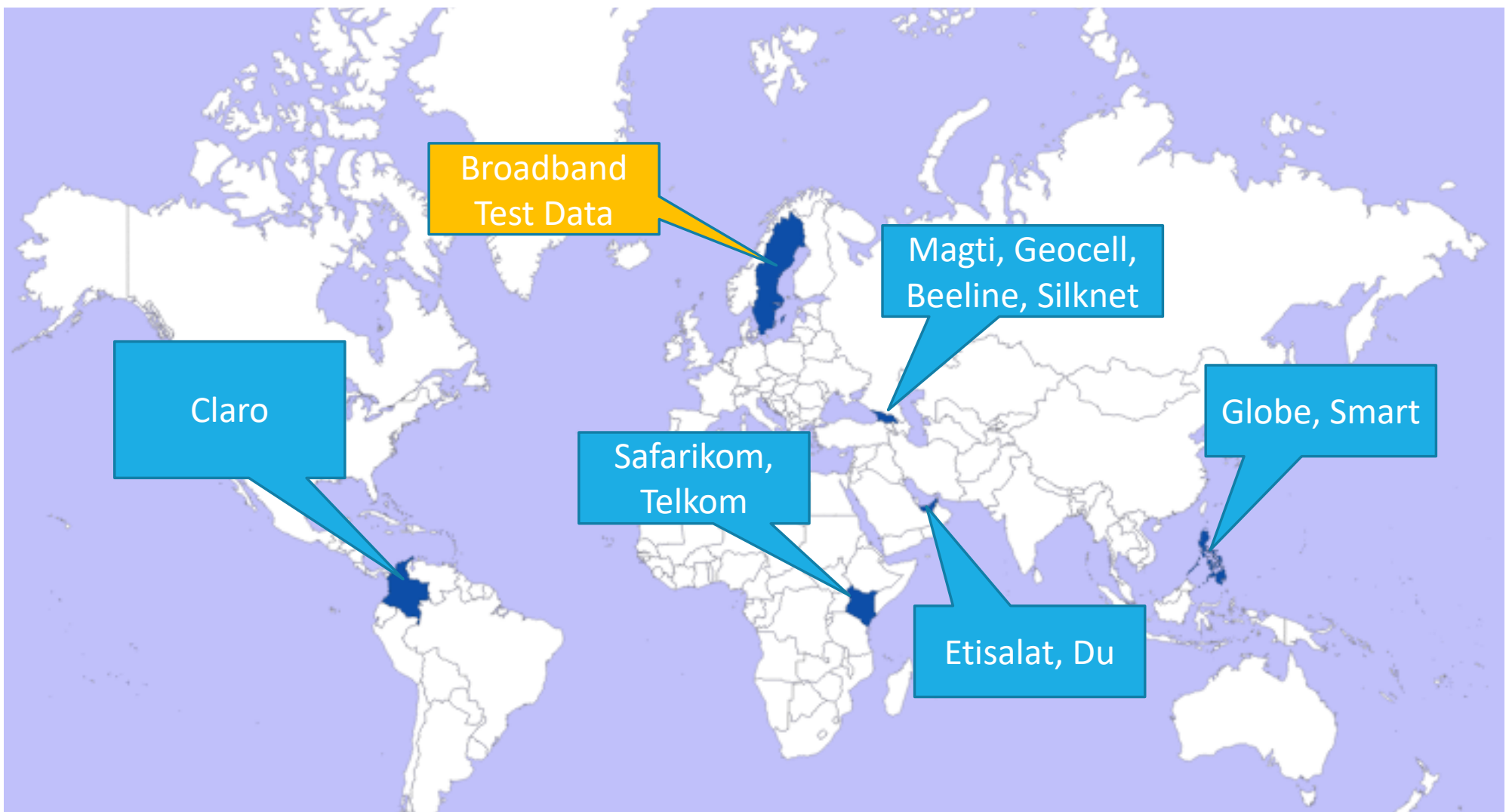




Objectives

- Complement existing indicators (granularity, disaggregation)
- New indicators

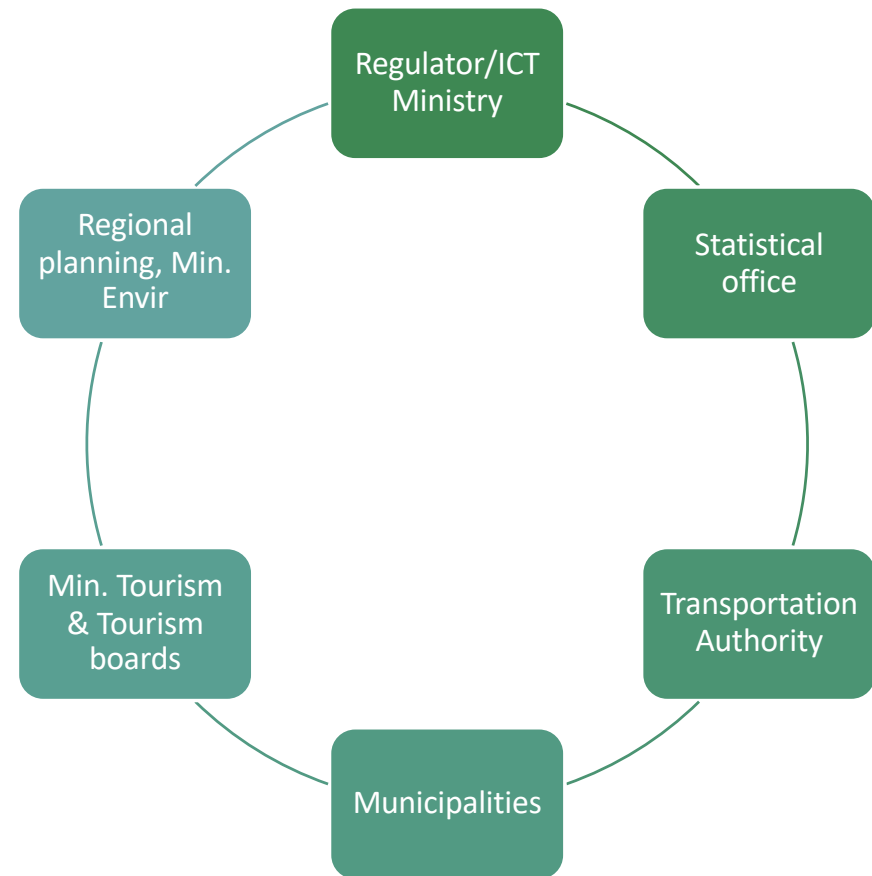
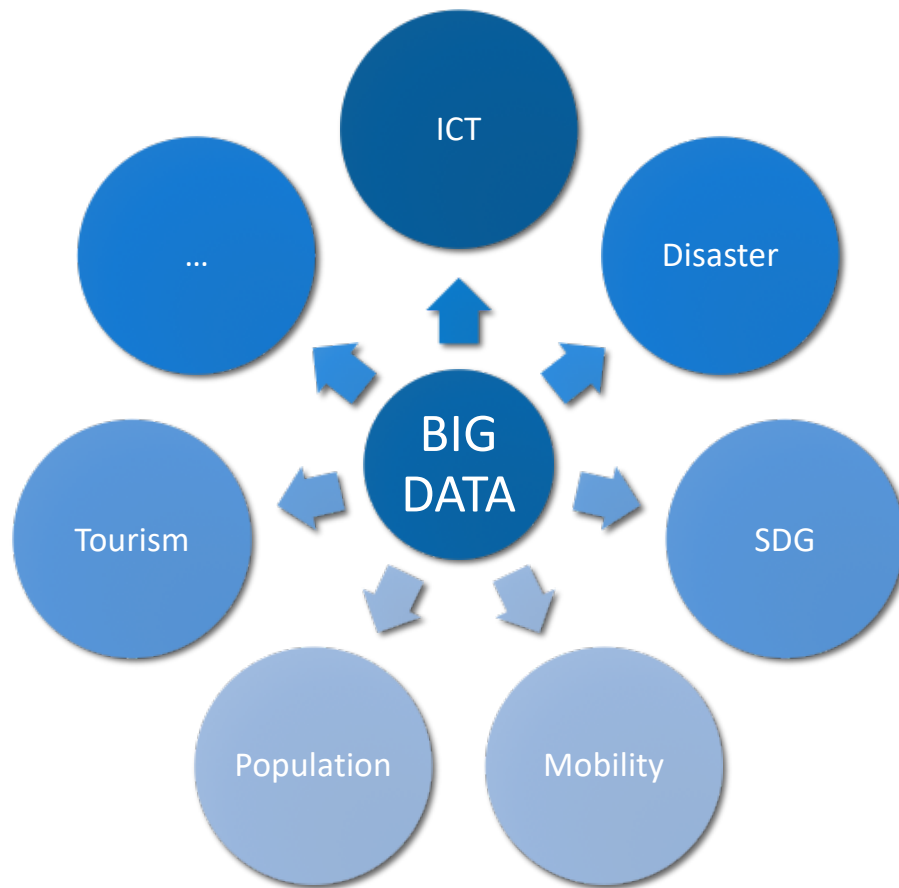




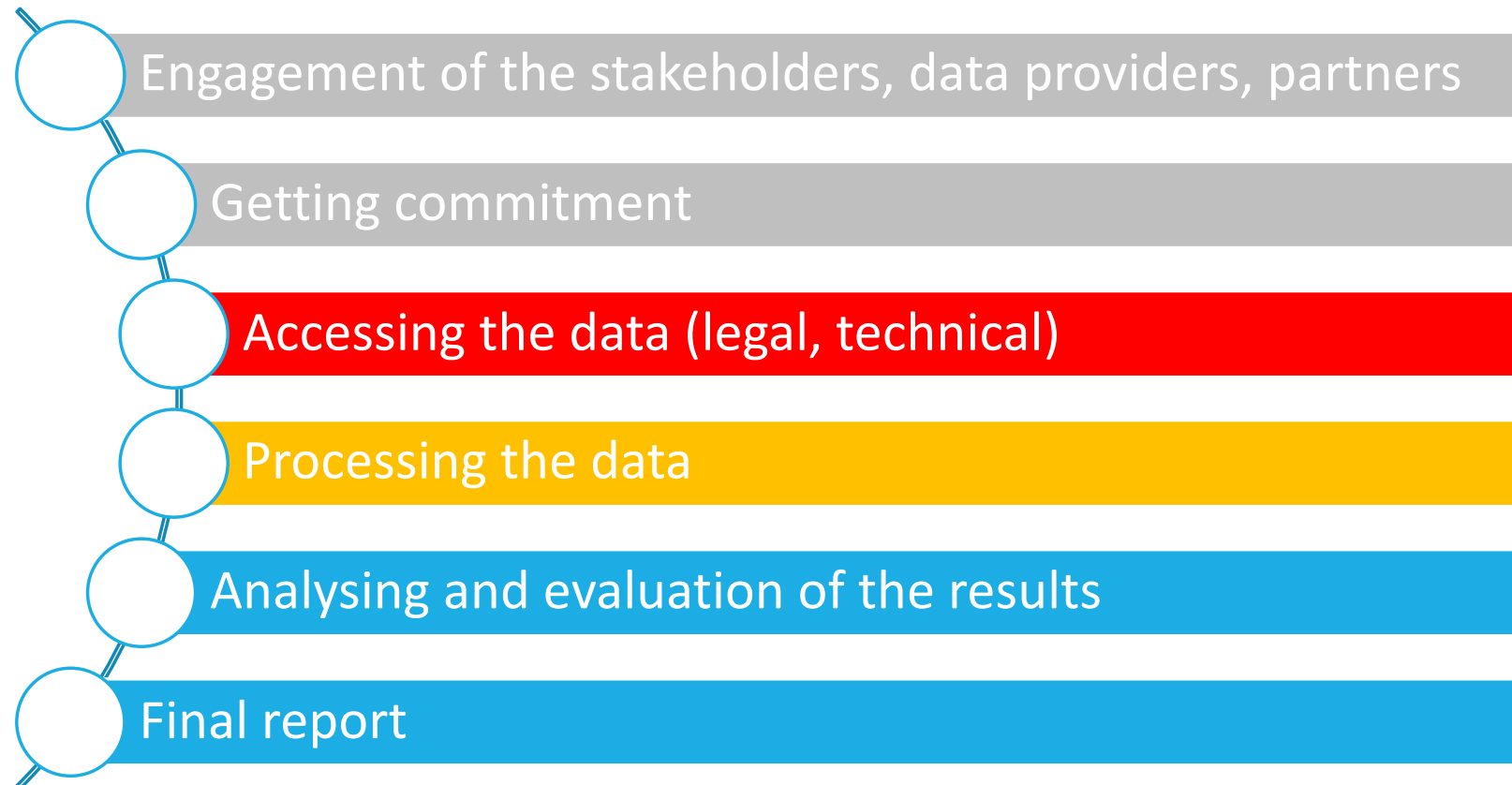
6 Pilot Countries

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

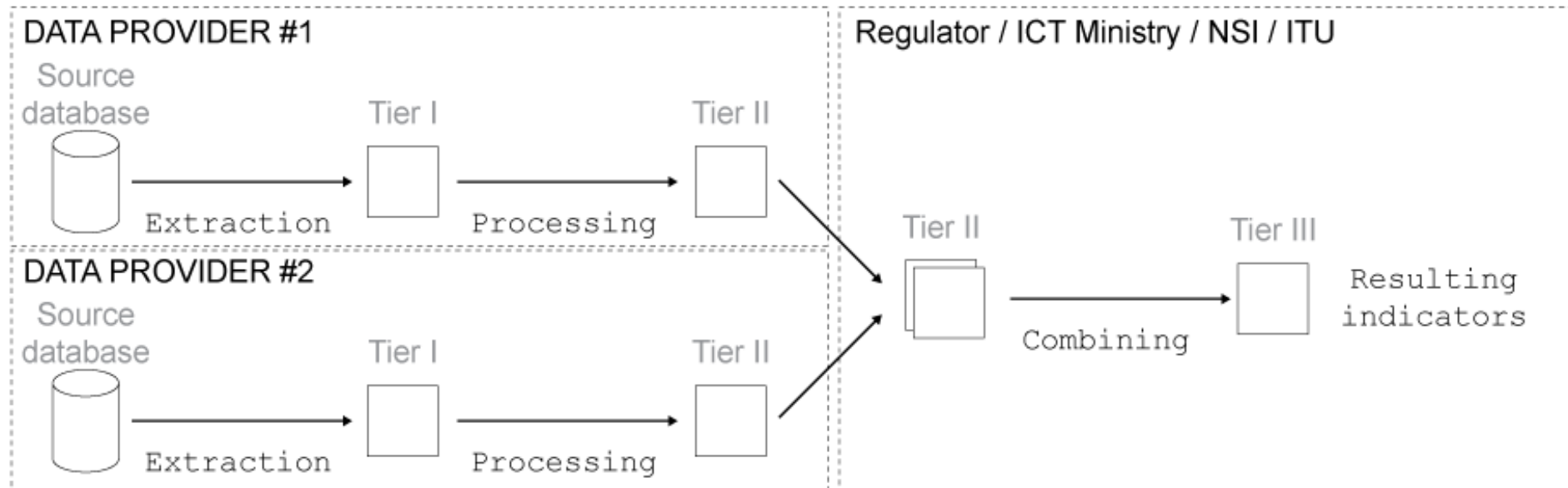
Stakeholders & data use



June 2016 – November 2017

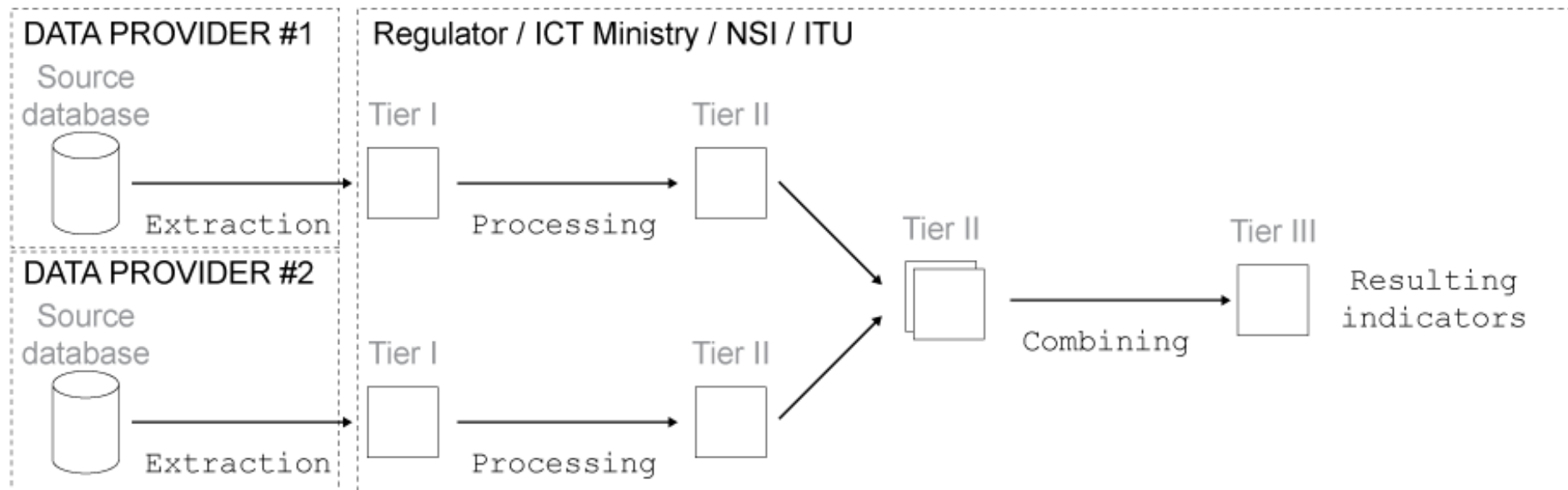


Data processing model – Option 1



- Kenya, Philippines and the UAE

Data processing model – Option 2



- Georgia, Colombia and Sweden



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 Speed, 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+: any proposed indicators from the country stakeholders

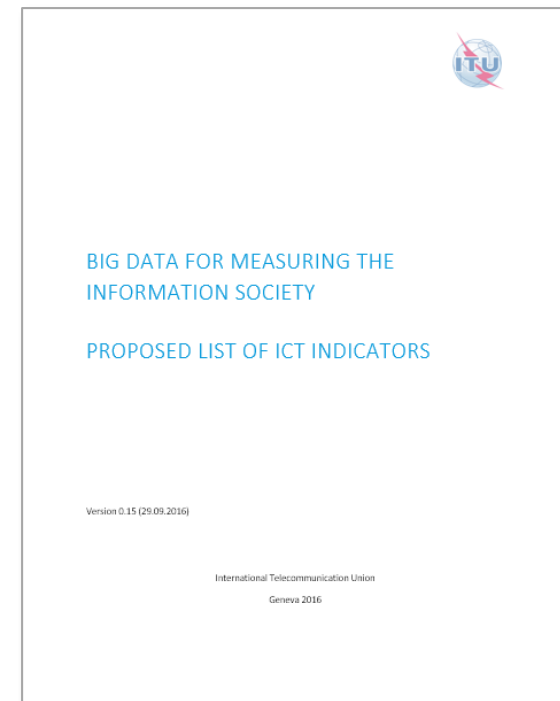
	Colombia	Georgia	Kenya	Philippines	Sweden	UAE
BD01	-	-	+	-	-	-
BD02	-	+	+	-	-	-
BD03	+	+	+	+	-	+
BD04	+	+	-	+	-	+
BD05	+	+	+	+	-	+
BD06	+	+	-	+	-	+
BD07	+	+	+	+	-	-
BD08	-	+	+	+	-	+
BD09	-	+	+	+	-	+
BD10	-	-	-	-	-	-
BD11	+	+	+	+	-	+
BD12	-	+	-	-	-	+
BD13	+	+	+	+	-	-
BD14	-	+	-	-	-	+
BD15	-	+	-	-	-	+
BD16	++++	+	-	-	-	+
TOTAL	11	14	9	9	0	11



Methodology document



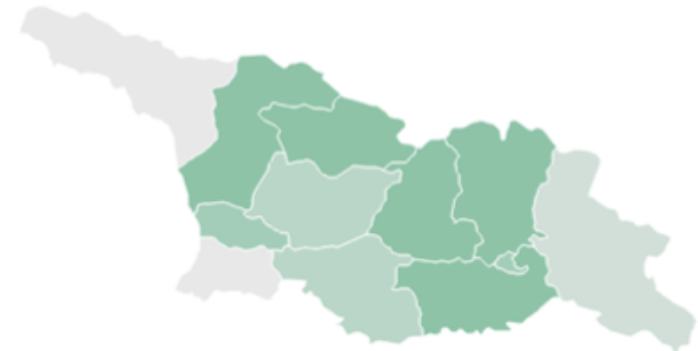
- ✓ Includes:
 - Name of the indicators
 - Data source description
 - Processing methodology
 - Expected results example
 - Disaggregation
 - The purpose and value of the indicator
- Complemented and amended during the project



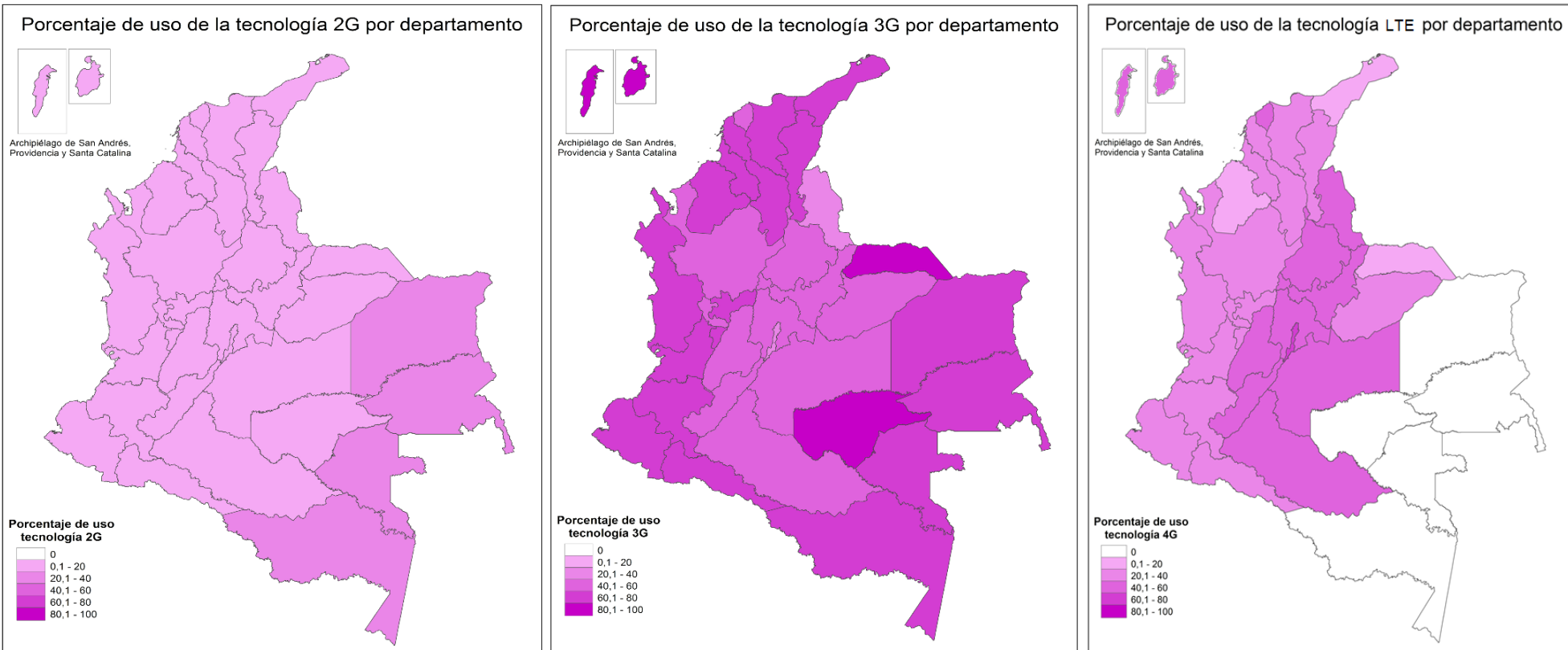


Difference with current reporting system

- Breakdown
 - Geographical: Local Administrative Units 1-2-3
 - Urban/Rural breakdown
 - Contract type: residential/non-residential; prepaid/postpaid; voice/data
 - Mobile technology generation: 2G, 3G, LTE
 - 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



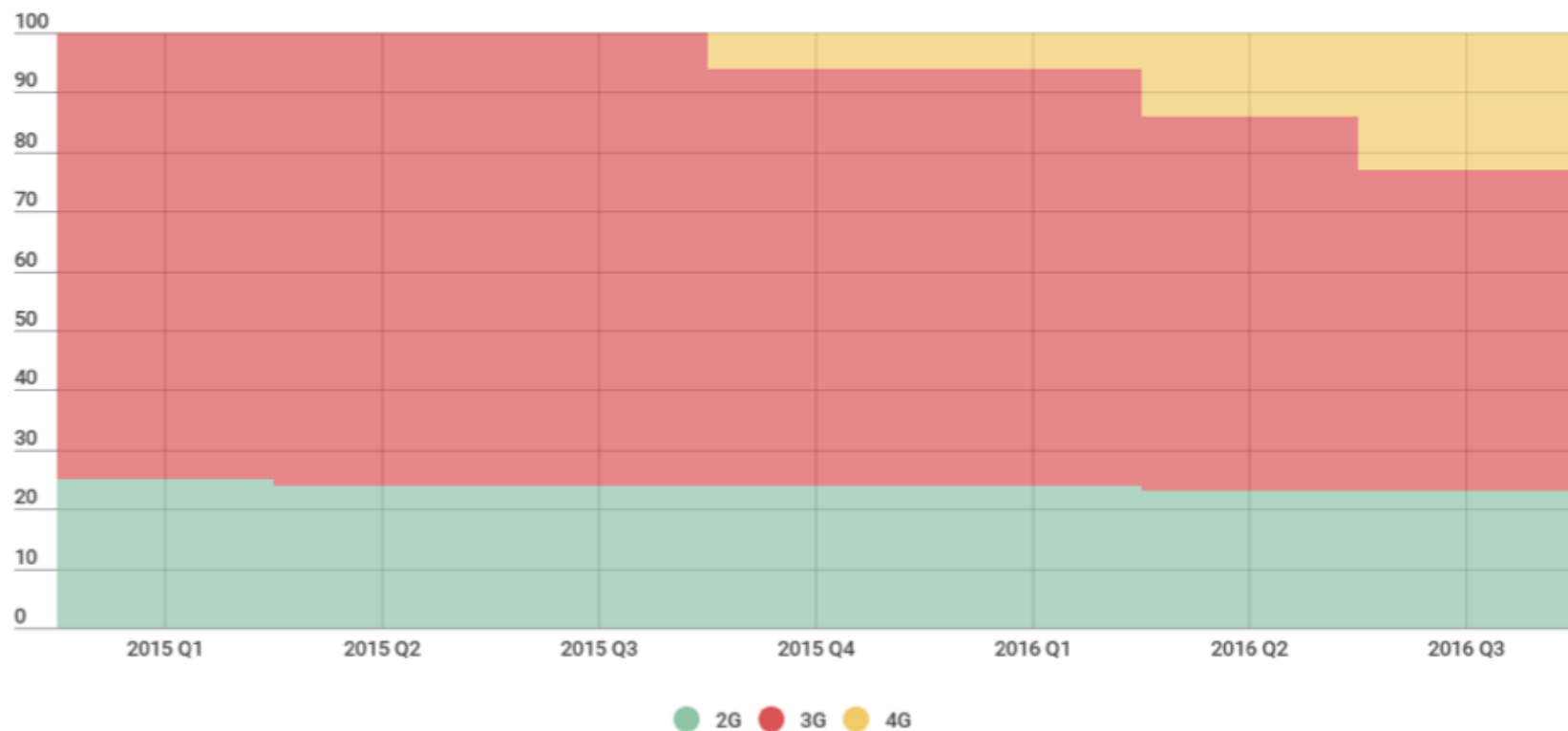
Geographical granularity





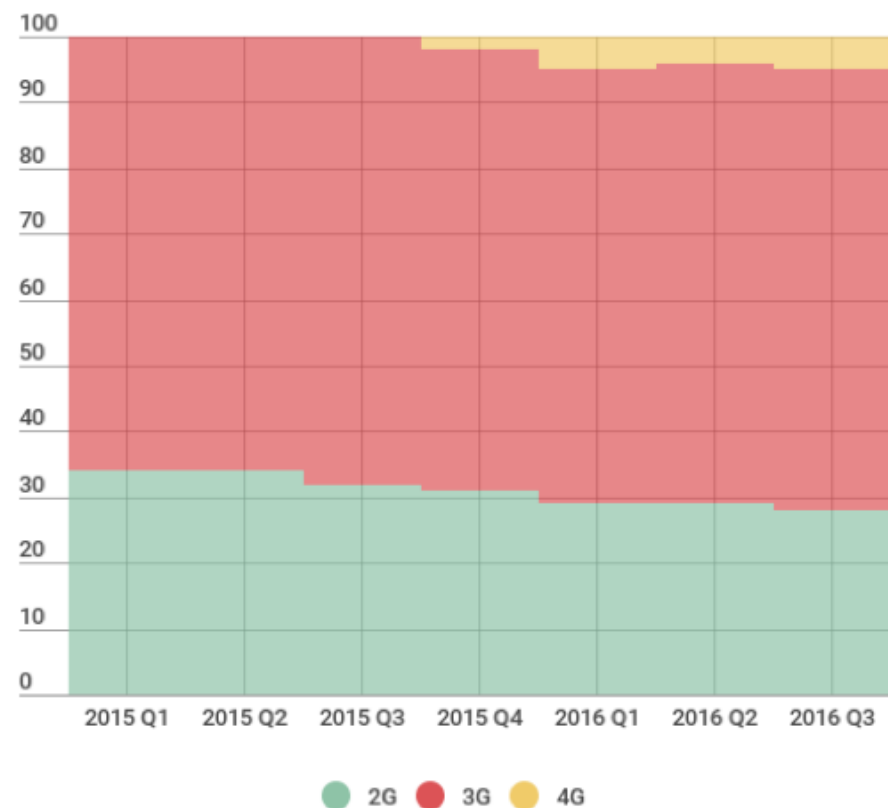
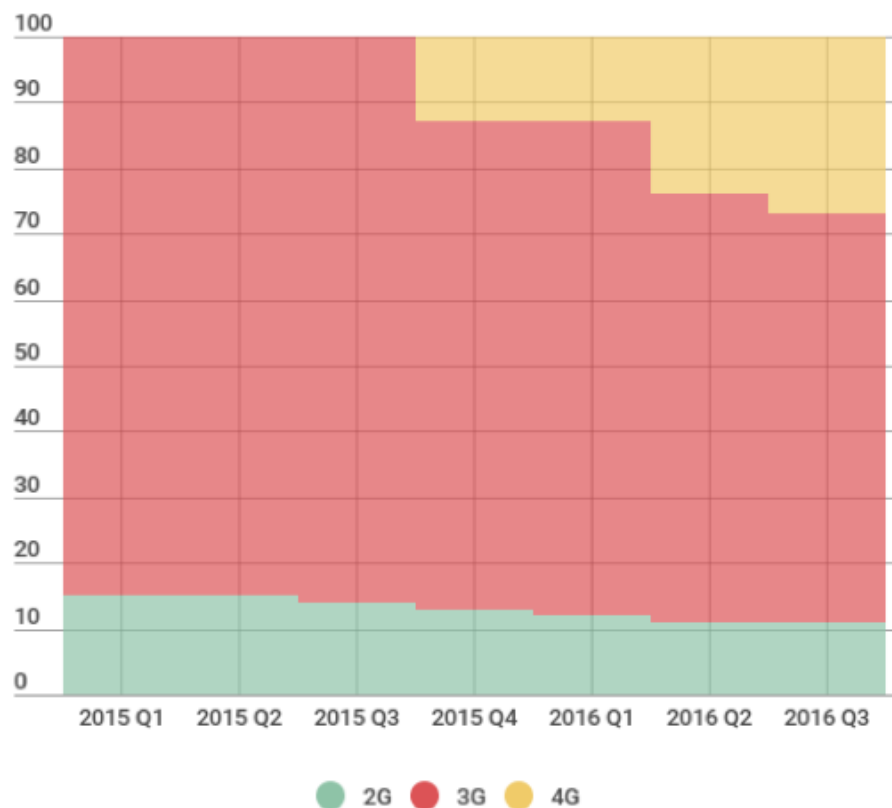
Breakdowns – by technology

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



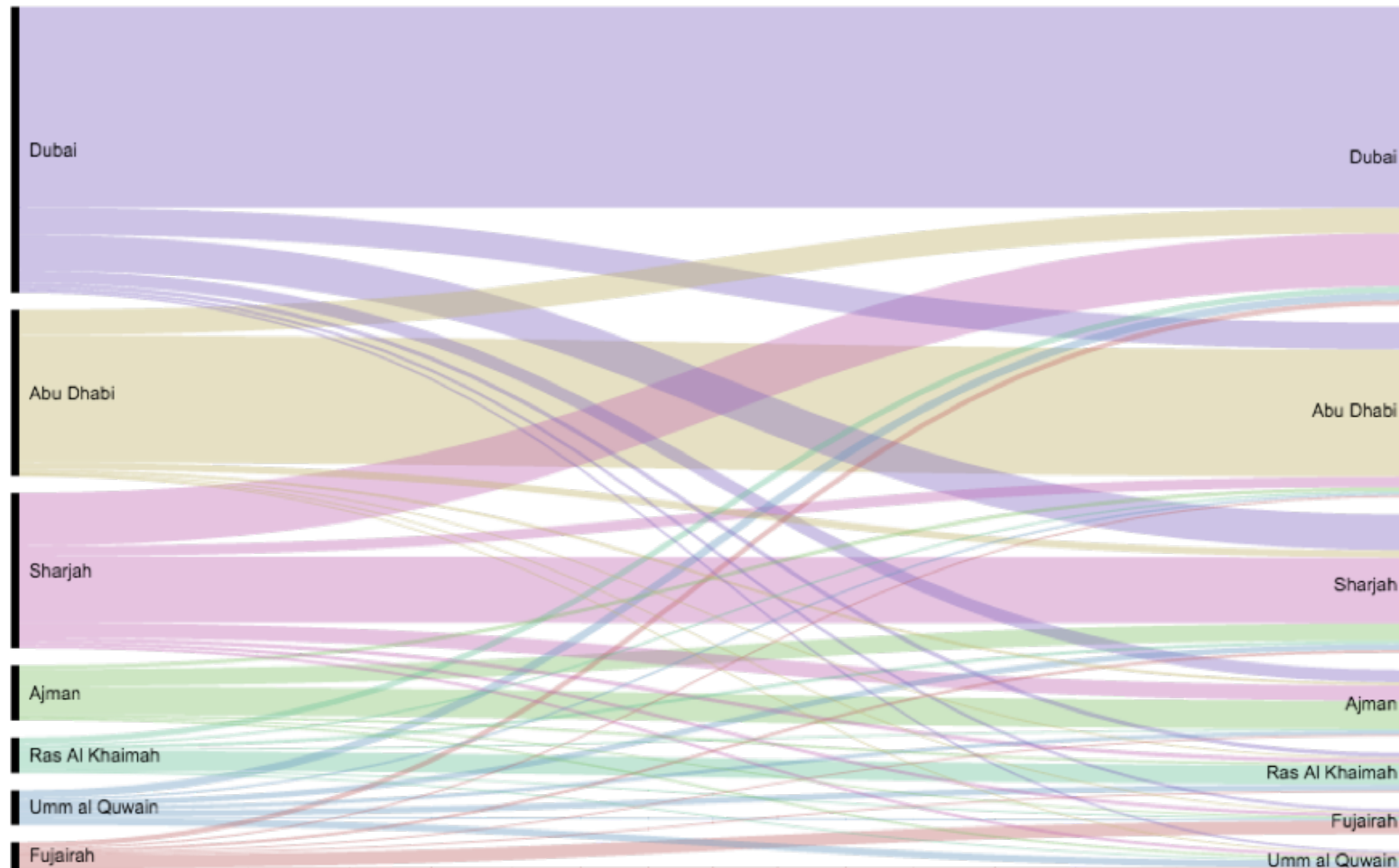


Breakdowns – rural/urban





New indicators



Origin/Destination Matrix for UAE administrative regions



Key issues faced

1. Access to data - administrative and legal procedures
2. Participation of all data providers (MNOs, ISPs)
3. Resources
 - Human (data scientists)
 - Infrastructure

Lessons Learned (for Phase II)



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)

Expertise required

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



<http://bigdata.black/technologies/data-science/how-to-become-a-data-scientist/>

Index

- Relevance of big data from the ICT sector
- Big data for measuring the information society
- **Experience in CIS countries – tour de table**



Big data in CIS countries

In your country:

1. Have you carried out any project to explore the use of big data from the ICT sector?
2. Are there plans to undertake projects in this area in the future?
3. Which are the main obstacles to using big data from the ICT sector for development and/or official statistics?





More information:

<http://www.itu.int/ict>
indicators@itu.int