BIG DATA FOR MEASURING THE INFORMATION SOCIETY: COUNTRY REPORT – UNITED ARAB EMIRATES

TABLE OF CONTENTS

Table of contents 1

List of tables and figures 3

- Tables 3
- Figures 3

Abbreviations used in the report 5

1. Background and Context 6
   - 1.1. Project Description 6
   - 1.2. Pilot Countrty Context 7
   - 1.3. Pilot Project Timeline 8

2. Stakeholders in the Project 10

3. Getting Access to the Data: Procedures, Legal Documents and Challenges 11
   - 3.1. Project Timeline 11
   - 3.2. Legal Documents and Challenges 11
   - 3.3. Resources Required for the Processing 13

4. Methodology, Technical Description, Tools 14
   - 4.1. Description of the Data Used in the Project 14
   - 4.2. Reference Data used in the Project 16
   - 4.3. Access to the Data 16
   - 4.4. Processing the Data 17

5. Results Derived for the Big Data Indicators 23
   - 5.1. BD03: Usage of Mobile-Cellular Networks for non-IP-Related Activities, by Technology 23
   - 5.2. BD04: Usage of Mobile-Cellular Networks for Internet Access, by Technology 26
   - 5.3. BD05: Number of Subscriptions with Access to Technology 29
   - 5.4. BD06: Active Mobile Voice and Broadband Subscriptions, by Contract Type 32

5.5. BD08: Active Mobile Devices 34
5.6. BD09: IMEI Conversion Rate 36
5.7. BD11: Mobile Domestic Broadband Traffic, by Contract Type, Technology........ 38
5.8. BD12: Mobile International Broadband Traffic, by Contract Type 38
5.9. BD14: Fixed Broadband Subscriptions, by Technology........ 39
5.10. BD15: Fixed Broadband Subscriptions, by Speed 40
5.11. BD16: Origin/Destination Matrix 41
6. Recommendations 44
7. Conclusions 47

LIST OF TABLES AND FIGURES

TABLES

Table 1: Key performance indicators on ICT development in the United Arab Emirates

Table 2: Project timeline

Table 3: Example of the template for the data request from the telecommunication service providers in the United Arab Emirates

Table 4: Example of raw data from MNO

Table 5: Example of raw data from ISP

Table 6: Local Administrative Unit code mapping table

FIGURES

Figure 1: Processing of the indicators

Figure 2: Snapshot of CSV file for a Tier II indicator

Figure 3: Snapshot of BD16 Tier II file

Figure 4: Data acquisition process for the United Arab Emirates

Figure 5: Distribution of technology for non-IP-related activities for the entire United Arab Emirates

Figure 6: Distribution of technology for non-IP-related activities disaggregated by Emirates

Figure 7: Distribution of mobile Internet access technology for the United Arab Emirates

Figure 8: Distribution of mobile Internet access technology according to administrative areas

Figure 9: Distribution of subscriptions for each technology type for the entire United Arab Emirates

Figure 10: Distribution of subscriptions for each technology type and administrative region

Figure 11: Number of active subscriptions by type for the entire United Arab Emirates

Figure 12: Number of contract types (data only, voice and data; prepaid, postpaid) for each administrative region

Figure 13: Number of active devices, by type, for the entire United Arab Emirates

Figure 14: Number of active devices, by type, for each administrative region

Figure 15: The percentage of IMEI change during the second quarter of 2016 for each administrative region

Figure 16: Traffic volume for mobile Internet activity for each administrative region, by technology

Figure 17: Traffic volume for mobile Internet generated by domestic subscribers in the United Arab Emirates and abroad

Figure 18: Fixed wired broadband subscriptions by technology, for each administrative region

Figure 19: Number of subscriptions according to speed, for each administrative region

Figure 20: Origin/destination matrix for the United Arab Emirates administrative regions, related to the second quarter of 2016

Figure 21: Flows of people between Abu Dhabi, Dubai and Sharjah for each month of the second quarter of 2016

ABBREVIATIONS USED IN THE REPORT

CDR	Call Detail Record
CSV	comma-separated values
FCSA	Federal Competitiveness and Statistics Authority
FTTH	fibre-to-the-home
ICT	information and communication technology
IMEI	International Mobile Station Equipment Identity
IPDR	Internet Protocol Data Record
ISP	Internet service provider
ITU	International Telecommunication Union
LAU	Local Administrative Unit
MNO	mobile network operator
NDA	non-disclosure agreement
TRA	Telecommunications Regulatory Authority

1. BACKGROUND AND CONTEXT

1.1. PROJECT DESCRIPTION

The aim of this project is to use big data from the telecom industry to improve and complement existing statistics and methodologies to measure the information society. The results of the project are expected to help countries and ITU to produce official information and communication technology (ICT) statistics and to develop new methodologies by combining new and existing data sources. This will directly benefit policy-makers, who will have access to new official statistics and benchmarks. It will also benefit data producers at the national level by guiding them in the use of big data for ICT measurement.

Specifically, the project:

- Combines the role of ITU as a standard-setting organization in terms of global ICT measurement with official data producers' experience and interest in working with big data. Those data producers who can and are willing to share big data are invited to join the project to explore new methodologies and data sources to produce official ICT statistics. ICT data producers include: national statistical offices, telecommunication regulatory authorities, telecommunication service providers (mobile network operators (MNOs)) and Internet service providers (ISPs).
- Takes advantage of ITU's involvement and experience in existing big data initiatives: ITU has organized a number of big data sessions in terms of regulation and monitoring, published information on the opportunities and challenges of big data, and developed a big data project on mobile phone data for health. ITU is also an active member of the United Nations Global Working Group on Big Data for Official Statistics, has networked with public and private organizations working on the topic, and has identified a number of potential (public and private) partners that would be interested in and benefit from participating in this project.
- Explores and analyses the kind of new information society data and statistics that ITU and other stakeholders (including policy-makers, analysts and other data producers and users) would value most, and which currently do not exist.
- Engages with ICT data producers to discuss the development of specific indicators for new statistics that could be produced through big data analytics, and by combing new and existing data sources and methodologies.

The first phase of the project involves several country pilot studies, which serve as concrete examples of how big data from the ICT industry (MNOs, ISPs, etc.) can be used and/or combined with existing official data (from ITU, national statistical offices and regulators) to produce new statistics, benchmarks and methodologies to measure the information society. These pilot studies cover different types of big data sources (e.g. MNOs and ISPs) and several ITU regions.

This report summarizes the pilot project conducted in the United Arab Emirates. More specifically, it highlights the roles, responsibilities and contributions of the different partners, the types of big data that were used, the big data analytics that were carried out, the results that were produced in terms of new indicators and statistics, and methodologies. It also provides recommendations on which and how data can be produced, collected, explored, aggregated, shared, analysed and presented. It discusses data access agreements and addresses issues and challenges related to possible data transfer, storage, privacy and confidentiality. The country reports are expected to provide data producers with concrete steps on using big data to produce new ICT statistics. An overall project report (forthcoming) will summarize the main findings from the pilot projects and make recommendations on how countries and ITU could use the results of the pilot studies for their ICT data collection.

1.2. PILOT COUNTRTY CONTEXT

The United Arab Emirates is the second-largest economy in the Arab world and is highly innovative in the ICT industry, due to the large and fast investments in ICT infrastructure seeking alternative economic growth to the petroleum industry. The United Arab Emirates has a well-developed and technologically advanced telecom sector with high levels of mobile, fixed-telephone and broadband penetration. Key factors in the growth of the United Arab Emirates telecom sector are robust economic growth, foreign investment and government initiatives. The United Arab Emirates has recently attracted a high number of ICT companies to host their regional headquarters with the latest communications infrastructure to service them. Ranked 40th in the 2017 ITU ICT Development Index (24th in the access sub index), the United Arab Emirates has shown great interest in international ICT and big data initiatives.

Telecommunications in the United Arab Emirates is under the supervision of the Telecommunications Regulatory Authority (TRA), with two main telecommunication providers: Emirates Telecommunications Group Company (Etisalat) and Emirates Integrated Telecommunications Company (EITC), represented as the brand “du”. There are 229.8 mobile subscriptions per 100 inhabitants in the United Arab Emirates.

The Federal Competitiveness and Statistics Authority (FCSA), a national statistics office, showed high interest in participating in this pilot project, as the United Arab Emirates is very interested in big data initiatives that would further advance the data-driven society and innovations in ICT. Key performance indicators showing the United Arab Emirates’ level of ICT development compared to the regional and global averages are presented in Table 1.

Table 1: Key performance indicators on ICT development in the United Arab Emirates

Indicator	2015	2016	Arab States	World
Fixed-telephone subscribers per 100 inhabitants	23.6	23.4	7.7	13.6
Mobile-cellular subscribers per 100 inhabitants	187.3	204.0	107.1	101.5
Fixed-broadband subscribers per 100 inhabitants	12.9	13.3	4.7	12.4
Active mobile-broadband subscribers per 100 inhabitants	130.9	156.7	45.2	52.2
3G coverage (% of population)	100.0	100.0	81.9	85.0
LTE/WiMAX coverage (% of population)	95.3	98.0	33.8	66.5
Mobile-cellular prices (% of GNI per capita)	0.2	0.2	2.5	5.1
Fixed-broadband prices (% of GNI per capita)	1.1	0.5	6.2	15.6
Mobile-broadband prices prepaid 500 MB (% of GNI per capita)	0.4	0.5	2.4	4.1
Mobile-broadband prices postpaid 1 GB (% of GNI per capita)	0.7	0.8	4.0	7.1
Percentage of households with computers	89.3	91.0	43.3	46.6
Percentage of households with Internet access	95.4	94.3	45.3	51.5
Percentage of individuals using the Internet	90.5	90.6	41.8	45.9
Internet bandwidth per Internet user (kbit/s)	108.8	133.7	39.0	74.5

Source: ITU.

1.3. PILOT PROJECT TIMELINE

The overall timeline of the pilot project, starting with the initial communication between ITU and FCSA, is shown in Table 2.

Table 2: Project timeline

Activity	Time
Initial communication, stakeholder involvement and project planning activities	June–October 2016
Kick-off meeting of the project with all stakeholders in the United Arab Emirates	End of October 2016
Preparations for processing the data by the telecommunication service providers, assessing the methodology for calculating the indicators and planning/allocating the resources for the tasks, resolving legal aspects	November 2016–January 2017
Precalculation of the indicators prior to the visit of the ITU data scientist	January 2017
ITU data scientist’s mission to the United Arab Emirates	29 January–9 February 2017
Assessing the resulting indicators, acquiring some corrections from the telecommunication service providers, initiating the country report, getting feedback, comments, assessment from local stakeholders in the report	March–April 2017
Country report, finalization of the project in the United Arab Emirates	May–September 2017

2. STAKEHOLDERS IN THE PROJECT

There are four stakeholders in the pilot project in the United Arab Emirates:

- Federal Competitiveness and Statistics Authority: The authority responsible for collecting national statistics. In the current project, FCSA is a country focal point responsible for coordinating the activities during the project in the United Arab Emirates.
- Telecommunications Regulatory Authority: The official telecommunication regulator and government body responsible for collecting ICT statistics from communication providers. TRA responsibility in the project is to oversee and manage the data collection from the telecommunication service providers.
- Etisalat–Emirates Telecommunications Group Company: The telecommunication service provider in the United Arab Emirates.
- du–Emirates Integrated Telecommunications Company: The telecommunication service provider in the United Arab Emirates.

In the pilot project, FCSA was the focal point and liaison for the communications between ITU and local stakeholders. FCSA's responsibility was to coordinate the activities between the stakeholders, facilitate general meetings and supervise the exchange of necessary documents and information. FCSA was also the main instigator of the pilot project in the United Arab Emirates.

TRA, the telecommunication regulator and main collector of ICT statistics, is responsible for reporting the ICT statistics to ITU. TRA was also the main organization responsible for facilitating the communications and workflow with the telecommunication service providers during the project. In the United Arab Emirates, there is no specialized data protection authority. Instead, TRA is the organization responsible for privacy and data protection aspects. This was also the case in the pilot project.

Etisalat and du are two main telecommunication providers in the United Arab Emirates responsible for providing the data for the project. Both telecommunication service providers have big data departments under their IT departments that were very capable of conducting the phase of the pilot involving their own data.

3. GETTING ACCESS TO THE DATA: PROCEDURES, LEGAL DOCUMENTS AND CHALLENGES

3.1. PROJECT TIMELINE

May 2016: The project started with confirmation of the local stakeholders about the willingness to participate in the project. This was initiated by FCSA and TRA, which approached service providers with a proposal to participate in the project.

June 2016: ITU sent an official project invitation letter to the TRA Director General, so TRA could participate officially in this pilot project.

July–August 2016: Discussion on planning the project, inviting telecommunication service providers, agreeing on kick-off meeting.

August–October 2016: Sending initial project documents, methodology description, work plan and agenda.

October 2016: Kick-off meeting in Abu Dhabi with all stakeholders, discussing available data, required legal documents and planning the visit of the data scientist. Creating template for data request from TRA to the telecommunication service providers.

November–December 2016: Identification of available data and allocating resources by the telecommunication service providers to conduct the project.

January 2017: Calculating some of the indicators by the telecommunication service providers.

January–February 2017: ITU data scientist's visit to the United Arab Emirates. Signing of non-disclosure agreements (NDAs) with TRA and two telecommunication service providers. Data scientist's role to support and provide assistance with methodology and any arising questions.

March–May 2017: Collecting data from the telecommunication service providers, analysis of the data, creating country report.

3.2. LEGAL DOCUMENTS AND CHALLENGES

To conduct the project, the following documents were necessary:

- Official invitation letter from ITU to TRA Director General to participate in the project;
- Official invination letter from TRA to du and Etisalat to participate in the project;
- Request of information from TRA to du and Etisalat;
- NDA between the data scientist (as a person, not representing ITU) and TRA;
- NDA between the data scientist (as a person, not representing ITU) and Etisalat;
- NDA between the data scientist (as a person, not representing ITU) and du.

The positive attitude by all stakeholders towards the project was crucial to enable a smooth process and swiftly overcome challenges encountered in different steps of the project cycle.

The main challenges in the project were administrative and legal issues, because the data scientist was acting as a private consultant. One of these challenges was identifying the legal options for the telecommunication service providers to provide the required data for the project, signing of NDAs between different parties, technical questions concerning processing and transmission of the data to ITU, and planning the resources of the telecommunication service providers to prepare the necessary data. However, these questions were successfully resolved.

Nevertheless, some legal issues took time and effort to resolve, such as identifying the necessary legal documents and specific procedures that FCSA and TRA had to take in order to officially request the participation and request of the data for the project. TRA had to make an official request for the telecommunication service providers to participate in the project.

One of the practical challenges was the planning of the kick-off meeting and visit of the data scientist. Due to the involvement of several countries in the project, the planning of these travels influenced the initial time plan of the project.

Another key challenge was identifying and signing the NDAs between different parties to assure the confidentiality of the data. Current practice in the United Arab Emirates is that NDAs are signed between organizations. Therefore, initially the NDA had to be signed by ITU and both telecommunication service providers; however, for practical reasons, this was not feasible (official signature by ITU would have taken too much time). Therefore, NDAs had to be signed directly between the telecommunication service providers and the data scientist, who was responsible for the processing and aggregation of the data.

As a result, the data scientist was required to sign three different NDAs with different stakeholders of the project. The NDAs were required, as the data scientist would have access to sensitive commercial information about the telecommunication service providers during the project. An issue, however, was that in the NDAs, the data scientist was acting as a private person and not a representative of ITU.

Concerning the data collection, a standard TRA data request template (in Excel) used for data collection of existing statistical indicators was adjusted to specify the data that were needed for the project from Etisalat and du. Both telecommunication service providers assessed the methodology and the template provided in order to map the specific indicators that could be calculated by both telecommunication service providers.

Table 3: Example of the template for the data request from the telecommunication service providers in the United Arab Emirates

	BD06: ACTIVE MOBILE VOICE AND BROADBAND SUBSCRIPTIONS, BY CONTRACT TYPE							
	Number of active mobile subscriptions disaggregated by the type of the contract (commercial, private; pre-paid, post-paid; no-data, voice and data, data only)							
LAU_LEVEL	LAU_CODE	TYPE_COMMERCIAL_PRIVATE (1-2)	TYPE_PRE_PST (1-2)	TYPE_VOICE_DATA (1-3)	GENDER (M/F/0)	AGE (year or 0)	LANG (EN, ... or 0)	COUNT
0	0	1	1	1				[illegible]
0	0	1	1	2				[illegible]
0	0	1	1	3				[illegible]
0	0	1	2	1				[illegible]
0	0	1	2	2				[illegible]
0	0	1	2	3				[illegible]
0	0	2	2	1				[illegible]
0	0	2	2	2				[illegible]
0	0	2	2	3				[illegible]
1	1	1	1	1				[illegible]
1	1	1	1	2				[illegible]
1	1	1	1	3				[illegible]
1	1	1	2	1				[illegible]
1	1	1	2	2				[illegible]
1	1	1	2	3				[illegible]
1	1	2	2	1				[illegible]
1	1	2	2	2				[illegible]
1	1	2	2	3				[illegible]
1	2	1	1	1				[illegible]
1	2	1	1	2				[illegible]
1	2	1	1	3				[illegible]
1	2	1	2	1				[illegible]
1	2	1	2	2				[illegible]
1	2	1	2	3				[illegible]
1	2	2	2	1				[illegible]
1	2	2	2	2				[illegible]
1	2	2	2	3				[illegible]

3.3. RESOURCES REQUIRED FOR THE PROCESSING

Private companies usually cannot quickly reallocate their resources to external, non-commercial projects (such as this project); however, both Etisalat and du were able to plan and reallocate the resources without any implications to the general project milestones. This was possible because both telecommunication service providers assigned special departments dealing with big data that were quickly able to identify the requirements for the project and plan the resources for that. The responsibility of these big data departments includes producing indicators similar to those requested. The departments have actual projects that require such outcomes and, as such, this project was in line with their everyday work.

For the overall processing of the data, the following human and technical resources were used:

- For calculating the indicators, there were four teams from all stakeholders with a minimum of two specialists from each and a back-office team working in parallel for a period of 30 working days. For both telecommunication service providers, a team of two to three persons was involved in the calculation process, for TRA and FCSA; also, two to three persons each were actively participating in the process.
- Teradata, Informatica, IBM Cognos, IBM Datastage and IBM Netezza Data warehouse platforms were used by the telecommunication service providers to extract, process and aggregate the

data for the project. These are the tools designed to manage and query big data, and are commonly used in the telecommunication industry for handling such data.

4. METHODOLOGY, TECHNICAL DESCRIPTION, TOOLS

4.1. DESCRIPTION OF THE DATA USED IN THE PROJECT

For the project, the following initial raw data were expected from the telecommunication service providers:

- The telecommunication service provider's geographical representation of antennas location or coverage areas. These data represent the geographical location of the Call Detail Records (CDRs) and Internet Protocol Detail Records (IPDRs) in the network, and are also used to assess the coverage area over the land and population.
- MNO's network usage data log records as CDRs and IPDRs. These records represent domestic, inbound roaming and outbound roaming data. These data were combined with available Customer Relationship Management data (contract types, demographics), and include all records individually for incoming and outgoing voice CDRs, incoming and outgoing messaging CDRs (SMS, MMS), and Internet traffic detail records (IPDR). Each event (usage log) has a separate individual record (e.g. one outgoing call is one record with attributes). Concerning the IPDR, the access to the Internet includes access to the broader public Internet (open Internet) and walled garden (closed Internet, limited access to walled garden content), but without distinction of the type of access. The observation period of the data was three months (April, May and June 2016).
- ISPs' fixed (wired)-broadband subscribers' Internet access log records.

CDRs are a well-known and used mostly standard format where each record represents an activity (i.e. call, SMS, Internet data usage session) performed by the subscriber, with attributes such as the time stamp of the beginning of the event, the antenna to which the device was connected, the duration of the activity, and some other features, tracked for operational and/or billing purposes.

Table 4: Example of raw data from MNO

ID	GENDER	AGE	LANG	IMEI	TYPE_COMMERCIAL_PRIVATE	TYPE_PRE_PST	TYPE_VOICE_DATA	TYPE_EVENT	DATETIME	DURATION	TYPE_DOMESTIC_OUTBOUND_INBOUND	MCC	CELL_ID	DATA_VOLUME
90000001				[illegible]	1	1	1	2	1463619156	96	1	209	[illegible]	
90000001				[illegible]	1	1	1	1	1463706028	10	1	209	[illegible]	
90000001				[illegible]	1	1	1	1	1463961322	504	1	209	[illegible]	
90000001				[illegible]	1	1	1	2	1464067148	229	1	209	[illegible]	
90000001				[illegible]	1	1	1	1	1464394213	302	1	209	[illegible]	
90000002				[illegible]	1	1	2	2	1461236210	1247	1	209	[illegible]	
90000002				[illegible]	1	1	2	2	1461238281	933	1	209	[illegible]	
90000002				[illegible]	1	1	2	4	1461447479	901	1	209	[illegible]	
90000002				[illegible]	1	1	2	6	1461701712	625	1	209	[illegible]	[illegible]
90000002				[illegible]	1	1	2	7	1461773942	132	1	209	[illegible]	[illegible]
90000003				[illegible]	1	1	3	6	1460985657	1359	1	209	[illegible]	[illegible]
90000003				[illegible]	1	1	3	7	1461179836	1451	1	209	[illegible]	[illegible]
90000003				[illegible]	1	1	3	6	1461363660	456	1	209	[illegible]	[illegible]
90000003				[illegible]	1	1	3	7	1461704553	1352	1	209	[illegible]	[illegible]
90000003				[illegible]	1	1	3	6	1461979761	26	1	209	[illegible]	[illegible]
90000004	M	35	EN	[illegible]	1	2	2	5	1462215001	736	1	209	[illegible]	
90000004	M	35	EN	[illegible]	1	2	2	1	1462419902	1400	1	209	[illegible]	
90000004	M	35	EN	[illegible]	1	2	2	6	1462651811	1630	1	209	[illegible]	[illegible]
90000004	M	35	EN	[illegible]	1	2	2	2	1462654988	1457	1	209	[illegible]	
90000004	M	35	EN	[illegible]	1	2	2	4	1462758074	336	1	209	[illegible]	
90000004	M	35	EN	[illegible]	1	2	2	5	1462947150	457	3	209		
90000004	M	35	EN	[illegible]	1	2	2	6	1463248854	1093	3	209		[illegible]
90000005	F	22		[illegible]	1	2	3	6	1461623716	1077	1	209	[illegible]	[illegible]
90000005	F	22		[illegible]	1	2	3	6	1461778309	192	1	209	[illegible]	[illegible]
90000005	F	22		[illegible]	1	2	3	6	1462036940	1723	1	209	[illegible]	[illegible]
90000005	F	22		[illegible]	1	2	3	6	1462340914	1101	1	209	[illegible]	[illegible]
90000005	F	22		[illegible]	1	2	3	6	1462445085	1932	1	209	[illegible]	[illegible]

Table 5: Example of raw data from ISP

ID	TYPE_COMMERCIAL_PRIVATE	TYPE_TECHNOLOGY	TYPE_SPEED	TYPE_IP_ACCESS	LAU3_CODE	DATETIME	DURATION	DATA_VOLUME
50000001	1	1	1	1	636732	1460590789	21021	[illegible]
50000001	1	1	1	1	636732	1460624755	19544	[illegible]
50000001	1	1	1	2	636732	1460667621	52585	[illegible]
50000002	1	1	2	2	736283	1463600670	37146	[illegible]
50000002	1	1	2	2	736283	1463655957	6527	[illegible]
50000002	1	1	2	2	736283	1463670975	78445	[illegible]
50000003	1	1	3	1	226398	1463201560	30617	[illegible]
50000003	1	1	3	1	226398	1463256930	43324	[illegible]
50000003	1	1	3	2	226398	1463302871	60706	[illegible]
50000004	1	1	4	2	109399	1460986631	72621	[illegible]
50000004	1	1	4	1	109399	1461087020	62676	[illegible]
50000004	1	1	4	2	109399	1461150692	1057	[illegible]
50000005	1	2	1	2	860843	1463270886	76957	[illegible]
50000005	1	2	1	1	860843	1463380473	39007	[illegible]
50000005	1	2	1	1	860843	1463436321	29605	[illegible]
50000006	1	2	2	1	448844	1460148452	61626	[illegible]
50000006	1	2	2	1	448844	1460249825	8365	[illegible]
50000006	1	2	2	2	448844	1460271473	4632	[illegible]

Country data for the United Arab Emirates are composed of CDRs and IPDRs collected and stored by Etisalat and du, covering 100 per cent of United Arab Emirates subscribers. The period of the data was the second quarter of 2016 (April, May and June). For mobile-cellular data, all the following records were included:

- Incoming and outgoing calling and SMS message events;
- Internet usage session events.

All mobile-cellular data were referenced to a specific Emirate via antenna identification, to which the specific event was connected when initiated.

For ISP data, fixed (wired)-broadband subscribers' Internet access log records were used. Also, the end location of the connection points was referenced to a specific Emirate.

A total of more than 44 trillion events were used as initial raw data by both telecommunication service providers combined.

4.2. REFERENCE DATA USED IN THE PROJECT

Reference data are required to be able to fully conduct the project according to the methodology. The reference data are used to link the data from the telecommunication service providers to geographical location (for geographical breakdown), and to calculate some specific indicators. As reference data, during the project, the following data sources were required and used:

- Administrative borders of the United Arab Emirates subdivisions (Local Administrative Units (LAUs)) along with the urban–rural subdivisions to be able to provide geographic and rural–urban breakdown of the indicators. In the United Arab Emirates, there is only one level of administrative units – seven Emirates and no urban–rural subdivisions. For the telecommunication service providers, there was no need to provide the maps for the seven Emirates, as they were able to allocate the indicators themselves from their data. The only requirement was to agree on the coding of each Emirate.
- One square kilometre grid of population. These reference data were not available for the project at the time of execution.

Because some of the reference data were not available in the United Arab Emirates at the time of the pilot project, some of the indicators (namely BD01 and BD02) were not implemented.

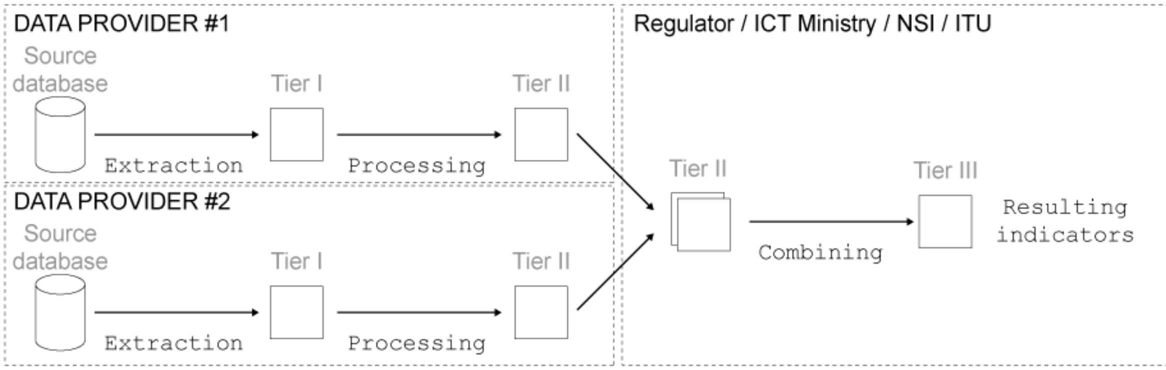
4.3. ACCESS TO THE DATA

Because of privacy protection reasons and because initial raw data from the telecommunication service providers can be considered to include confidential business information, there are three tiers or phases of the data processing:

- Tier I – Initial, raw, not aggregated data extracted by the telecommunication service providers from their databases and registries, that is the basis for calculation and which might include private and confidential business information. This can also be referred as microdata.
- Tier II – Initially aggregated data per telecommunication service provider with no private and some (or no) confidential business information, but which is still considered sensitive and kept from sharing with third parties. This can also be referred to as macrodata.
- Tier III – These Tier II aggregated indicators from different telecommunication service providers merged that can be publicly shared and do not include any private or confidential business information. Tier III represents the resulting statistical indicators of the project.

In the United Arab Emirates, the process of extraction and calculation of Tier II indicators was done on the premises of the telecommunication service providers and the Tier I microdata (raw data individual records) were not provided to ITU, TRA or any other party due to confidentiality issues (Figure 1).

Figure 1: Processing of the indicators



Both telecommunication service providers extracted the initial raw data from their data warehouses, relying on big data processing technologies. After the extraction, Etisalat and du processed the data according to the methodology and aggregated to Tier II data, which were stored as comma-separated values (CSV) files. The CSV files were delivered to the data scientist. Tier I data were not explicitly stored and the data scientist did not have access to them.

After Tier II data were produced by the telecommunication service providers, the procedure was to transmit the CSV files to TRA. TRA in turn then provided the files to the data scientist.

4.4. PROCESSING THE DATA

Both du and Etisalat have big data departments that are very competent for completing such calculations and, although not without some recalculation, both service providers were able to provide Tier II results to the data scientist successfully, who combined the data into Tier III resulting indicators.

The process of calculating Tier II aggregated data followed precise and well-defined steps. ITU, through TRA and FCSA, provided Etisalat and du the list of all the indicators required. Telecommunication service providers had allocated a team of data scientists in charge of accessing and extracting Tier I data, and processing them into Tier II data. The IT teams of both companies agreed on the interpretation of each indicator, and then provided Tier II data to the data scientist. The calculation of Tier II indicators by both telecommunication service providers was assisted by the data scientist, who helped with methodological questions where the help was necessary.

During the visits of the data scientist, the IT/big data departments of both telecommunication service providers agreed on the interpretation of the computing methodology for each of the indicators. These meetings were held together with the data scientist, who assisted and provided any necessary solution for issues raised by the telecommunication service providers. The data scientist did not have any access to Tier I data, nor to the algorithms and/or SQL queries to validate directly the extraction and processing of this phase.

Tier II calculations took place before, during and after the data scientist's mission to the United Arab Emirates. Most of the calculations were done before, some were calculated during, and one optional requested indicator was calculated after the mission.

Spatial aggregation of data has been made only by distinguishing between the whole country (LAU0) and main administrative regions (Emirates – LAU1), without any distinction between urban and rural areas. This decision is driven by the fact that second- (LAU2) and third-level (LAU3) administrative divisions are not available in the United Arab Emirates.

According to the list of indicators proposed by ITU, 10 of 15 Tier II indicators were computed by Etisalat and du. Two indicators (BD01 and BD02) were dropped due to the lack of data regarding land coverage area of antennas: such a feature is required to compute the dropped indicators. BD07, BD10 and BD13 were not calculated due to confidentiality aspects.

In addition to the proposed indicators, telecommunication service providers of the United Arab Emirates agreed to compute an additional indicator BD16: an origin/destination matrix, which is a well-known indicator to measure human mobility across a given territory, divided into regions/cities or any possible space aggregation. For the case of the United Arab Emirates, the origin/destination matrix has been computed among the seven Emirates of Abu Dhabi, Dubai, Sharjah, Ajman, Fujairah, Ras Al-Khaimah and Umm Al-Quwain.

The following indicators were calculated by both telecommunication service providers in the United Arab Emirates:

- 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;
- BD08: Active mobile devices;
- BD09: IMEI conversion rate;
- BD11: Mobile domestic broadband traffic, by contract type, technology;
- BD12: Mobile international broadband traffic, by contract type;
- BD14: Fixed broadband subscriptions, by technology;
- BD15: Fixed broadband subscriptions, by speed;
- BD16: Additional indicator: Origin/destination matrix.

Telecommunication service providers delivered Tier II indicators as standard CSV files, for the data scientist to aggregate the data into Tier III final indicators. Telecommunication service providers set up a CSV file for each indicator, resulting in a total of 20 CSV files and a separate two files for BD16 (origin/destination matrix indicators), composed by approximately 1 million rows/approximately 50 MB per indicator.

Figure 2: Snapshot of CSV file for a Tier II indicator

LAU_LEVEL	LAU_CODE	Date	TYPE_COMMERCIAL_PRIVATE	TYPE_PRE_PST	TYPE_VOICE_DATA	COUNT_UNIQUE
1	1	01/04/16	1	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	DATA Only	[illegible]
1	1	01/04/16	2	POSTPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	1	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	1	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	1	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	DATA Only	[illegible]
1	1	01/04/16	2	POSTPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	DATA Only	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	1	PREPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	VOICE and DATA	[illegible]
1	1	01/04/16	2	POSTPAID	DATA Only	[illegible]

Since BD16 was proposed during the visiting period of the data scientist, the BD16 indicator was provided at the end the visit. The specification of BD16 is provided as an annex of the list of indicators. The size of BD16 CSV files is smaller than other indicators (CSV file size <10 MB), as the aggregation level is at the country/Emirate level (LAU0/LAU1).

Figure 3: Snapshot of BD16 Tier II file

LAU3_CODE_MOST_FREQUENT_LOCATION	LAU3_CODE	TYPE_COMMERCIAL_PRIVATE	YEAR_MONTH	COUNT
Ajman	Sharjah	2	201605	[illegible]
Dubai	Ras Al Khaimah	1	201606	[illegible]
Dubai	Abu Dhabi	1	201604	[illegible]
Abu Dhabi	Dubai	2	201606	[illegible]
Ras Al Khaimah	Dubai	2	201606	[illegible]
Abu Dhabi	Ras Al Khaimah	1	201605	[illegible]
Abu Dhabi	Umm Al Qwaiwan	2	201606	[illegible]
Abu Dhabi	Fujairah	2	201606	[illegible]
Fujairah	Sharjah	1	201605	[illegible]
Umm Al Qwaiwan	Sharjah	2	201605	[illegible]
Abu Dhabi	Ras Al Khaimah	1	201604	[illegible]
Abu Dhabi	Sharjah	2	201604	[illegible]
Fujairah	Ajman	2	201606	[illegible]
Dubai	Sharjah	1	201605	[illegible]
Fujairah	Abu Dhabi	1	201606	[illegible]
Abu Dhabi	Fujairah	2	201605	[illegible]
Sharjah	Ras Al Khaimah	1	201605	[illegible]
Fujairah	Ras Al Khaimah	1	201605	[illegible]
Abu Dhabi	Ras Al Khaimah	1	201606	[illegible]
Sharjah	Dubai	2	201606	[illegible]
Sharjah	Umm Al Qwaiwan	2	201606	[illegible]
Fujairah	Dubai	2	201604	[illegible]
Dubai	Ras Al Khaimah	2	201604	[illegible]
Dubai	Ras Al Khaimah	2	201605	[illegible]
Dubai	Fujairah	2	201605	[illegible]
Abu Dhabi	Umm Al Qwaiwan	2	201604	[illegible]
Dubai	Abu Dhabi	2	201604	[illegible]
Ajman	Ras Al Khaimah	2	201606	[illegible]
Sharjah	Fujairah	1	201605	[illegible]
Abu Dhabi	Sharjah	1	201604	[illegible]
Abu Dhabi	Umm Al Qwaiwan	1	201606	[illegible]
Ras Al Khaimah	Abu Dhabi	1	201604	[illegible]
Ras Al Khaimah	Dubai	1	201606	[illegible]
Abu Dhabi	Fujairah	1	201605	[illegible]
Ajman	Umm Al Qwaiwan	1	201604	[illegible]
Umm Al Qwaiwan	Abu Dhabi	2	201605	[illegible]

Tier II data from both telecommunication service providers were validated according to the agreed methodology, but without the possibility for the data scientist to inspect codes and/or algorithms used in the aggregation from Tier I into Tier II data. After the initial check and refinement of some of the Tier II indicators, made by telecommunication service providers together with the data scientist, the final versions of each indicator's CSV files were sent to the data scientist. The refinements of Tier II CSV files were mostly to resolve formatting issues, such as normalizing the column names and number formats, in order for the CSV files from two telecommunication service providers to be merged. The Tier II data were stored in a local SQLITE3 database and aggregated into Tier III data.

Since access to Tier I data was not available, the validation on data has been made only at a logical basis, by inspecting values distributions and checking them with du and Etisalat big data teams.

There are two possible levels of validations:

- Algorithm validation along with Tier I data validation;
- Result validation.

As the data scientist was not able to access Tier I data nor inspect the algorithms that were used to extract, process and aggregate Tier I data into Tier II data, the data scientist could only validate the results, mostly using the logical approach to the resulting data.

Basically, the validation of the algorithms was done during the meetings with the telecommunication service providers, where the logic of queries was explained and agreed by both telecommunication service providers, along with supervision by the data scientist. The resulting data validation was done by the data scientist, mostly using the logical checks, to see if resulting Tier II indicators made sense and were comparable.

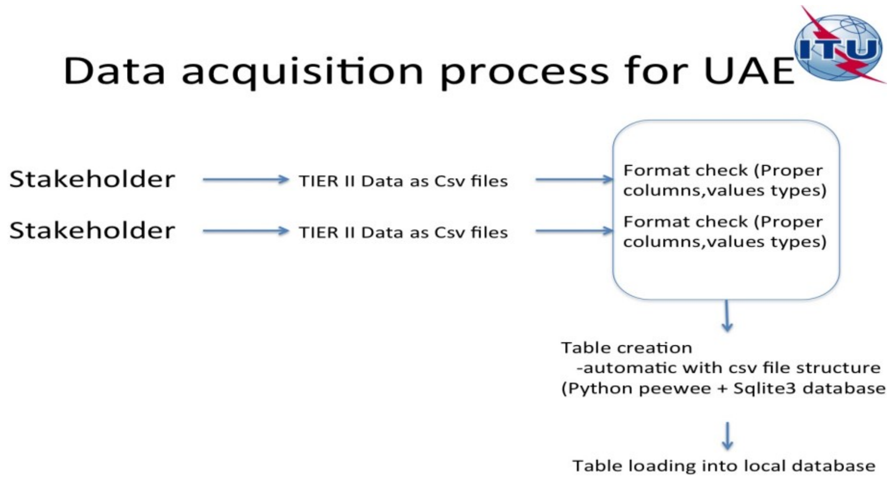
Checking the data format of Tier II has been the major challenge of the process: the provided CSV files were in a different format from each telecommunication service provider, with often different column names and different LAU1 codes conventions. Moreover, the separation character was different among CSV files, requiring a further editing of all the files provided (e.g. sometimes ";" was used as delimiter, sometimes ",", so the delimiters had to be changed to the same delimiter in all Tier II files). Given these required adjustments, the suggestion for future experiments is to develop an automatic tool to pre-process acquired data with a format check and a preliminary analysis in order to spot eventual "strange" outliers. For example, after the first acquisition, values for the Emirate of Abu Dhabi were different according to the telecommunication service providers. After a further check, it came out that they used different LAU1 codes, meaning the LAU1 codes used for Abu Dhabi were different from different telecommunication service providers. This had to be corrected, although the LAU1 codes were initially agreed upon in a mapping table (Table 6).

Table 6: Local Administrative Unit code mapping table

LAU_LEVEL	LAU_CODE	LAU1_NAME
0	0	United Arab Emirates
1	1	Ajman
1	2	Abu Dhabi
1	3	Dubai
1	4	Fujairah
1	5	Ras al-Khaimah
1	6	Sharjah
1	7	Umm al-Quwain

The merging of Tier II data into Tier III indicators and visualization has been performed using Jupyter Notebook and Bokeh plotting library, in order to obtain easy-to-export JavaScript plots, ready for an interactive web visualization. In Figure 4, the steps of the process of acquiring Tier II data from telecommunication service providers are depicted.

Figure 4: Data acquisition process for the United Arab Emirates



5. RESULTS DERIVED FOR THE BIG DATA INDICATORS

Indicators computed for the United Arab Emirates are 11 out of the 15 proposed by ITU. One additional indicator was also computed, namely BD16: Origin/destination matrix. It is a widely used indicator in the field of human mobility, and in recent years it is often based on CDR data. It represents the flow of persons across each possible pair of regions/territories.

5.1. BD03: USAGE OF MOBILE-CELLULAR NETWORKS FOR NON-IP-RELATED ACTIVITIES, BY TECHNOLOGY

This indicator provides insight on the actual usage of mobile technology for making mobile voice and messaging. This indicator is applied to MNOs. There is no associated official indicator corresponding to this indicator, so this can be considered a new ICT indicator.

The indicator should provide the understanding of actual usage of the mobile technology in different parts of the country for making and receiving mobile calls and messaging. The usage is defined as the number of connections made between the device and the network – e.g. one outgoing call in the 2G network is a fact of the usage of the 2G network, one incoming SMS via the 3G network is a fact of the usage of the 3G network. The resulting indicator should refer to the proportion of usage of the specific technology compared to the total usage of all generations of wireless mobile telecommunication technology (e.g. the usage proportion of 2G technology in the specific region divided by the total usage of all technologies in the specific region).

Figure 5: Distribution of technology for non-IP-related activities for the entire United Arab Emirates

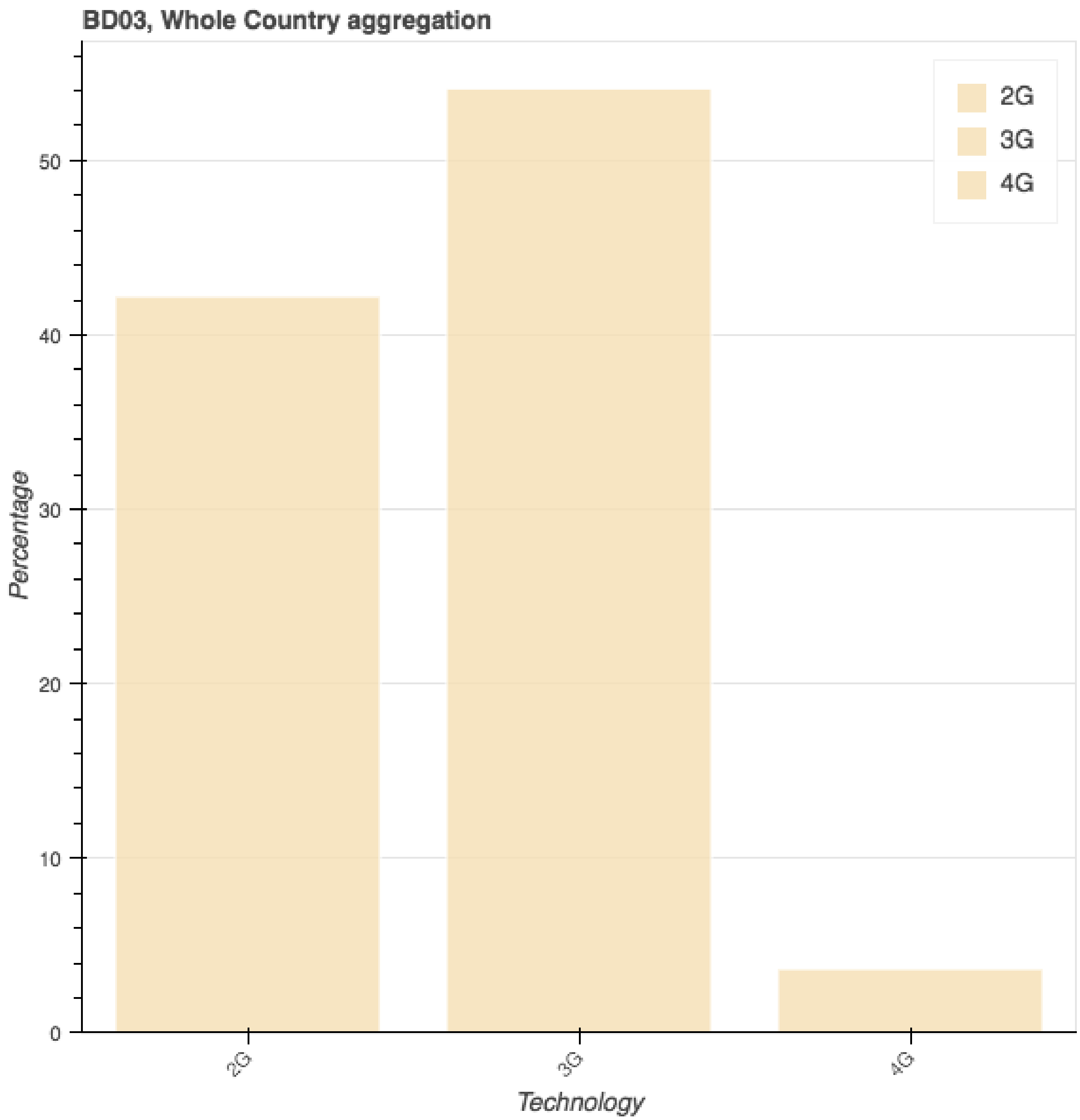
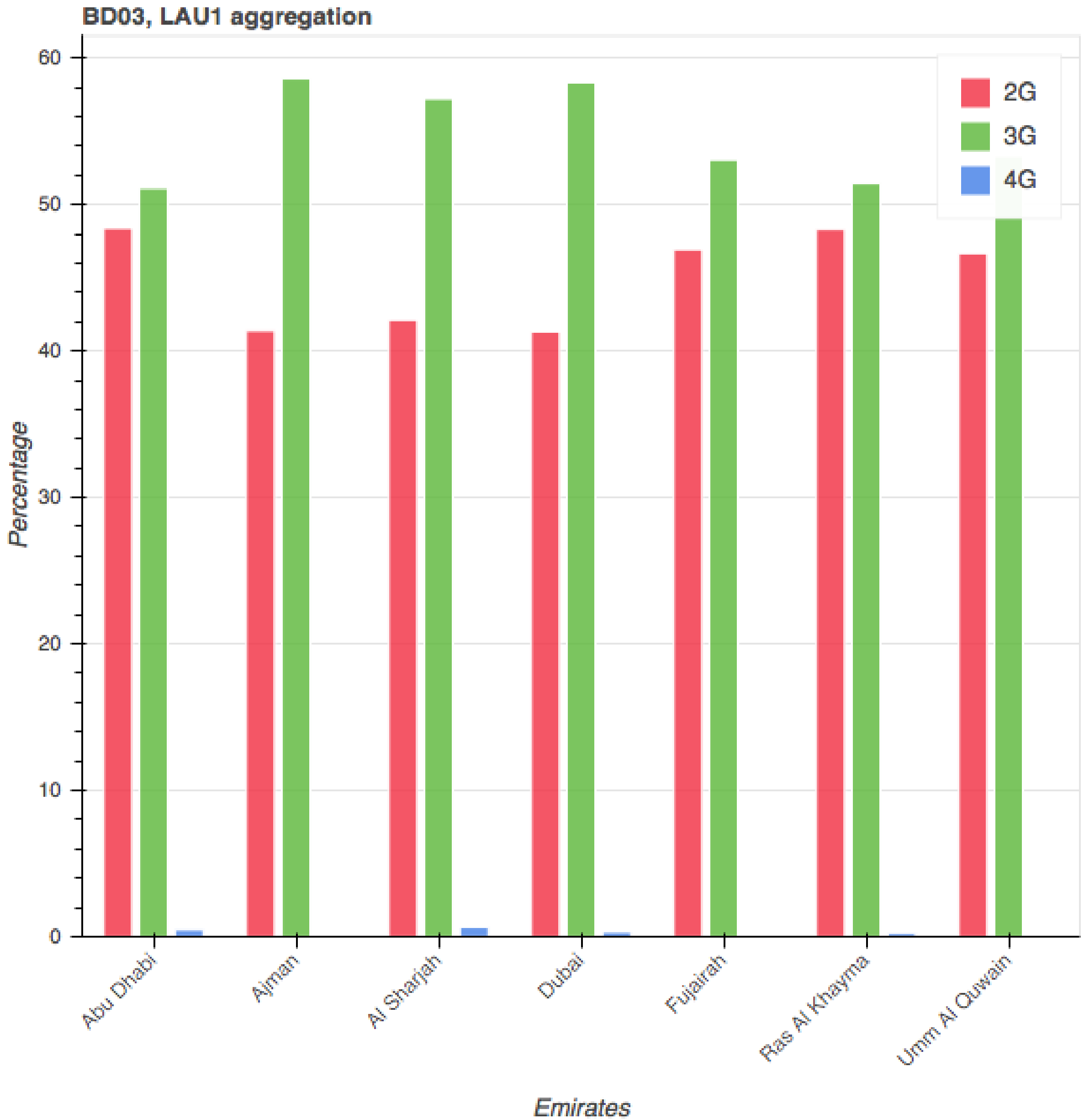


Figure 6: Distribution of technology for non-IP-related activities disaggregated by Emirates (LAU1 code)



Figures 5 and 6 show the use of the mobile-cellular network for non-IP activities (voice call, SMS, etc.), both for the entire United Arab Emirates and for each Emirate separately. The data clearly show how the traffic related to this indicator is not using the 4G network. This is because 4G is designed for IP traffic and not voice calls. The percentages for the usage of 2G and 3G are also rather similar among each of the Emirates, but minor differences can be observed in different Emirates. As seen from this indicator, the United Arab Emirates has successfully implemented a 3G network. There is no "perfect" division, but the trend over time should be to increase the usage of and access to higher technologies.

The regional differences in this indicator (as well as in combination with the following BD04) enable the telecommunication service providers and the Government of the United Arab Emirates to monitor the progress of development of the technologies and to support the decisions for investing in regions that are lagging behind.

5.2. BD04: USAGE OF MOBILE-CELLULAR NETWORKS FOR INTERNET ACCESS, BY TECHNOLOGY

This indicator provides the understanding of actual usage of the mobile technology for accessing the Internet in different parts of the country. The usage is defined as the number of connections made between the device and the network – e.g. one access to the Internet via the 2G network is a fact of the usage of the 2G network, one access to the Internet via the 4G network is a fact of the usage of the 4G network.

Figure 7 shows the distribution of mobile Internet access by technology for the whole country of the United Arab Emirates. The data clearly show that 3G and 4G are the standard for the mobile Internet in the United Arab Emirates, with 2G representing less than 10 per cent of connections. The indicator for the whole country shows that usage of the 4G network is very close to the usage of the 3G network in terms of connected sessions. There is no ideal breakdown by technology, but this indicator can provide a very good insight on the status of the country when comparing to other countries.

Figure 7: Distribution of mobile Internet access technology for the United Arab Emirates

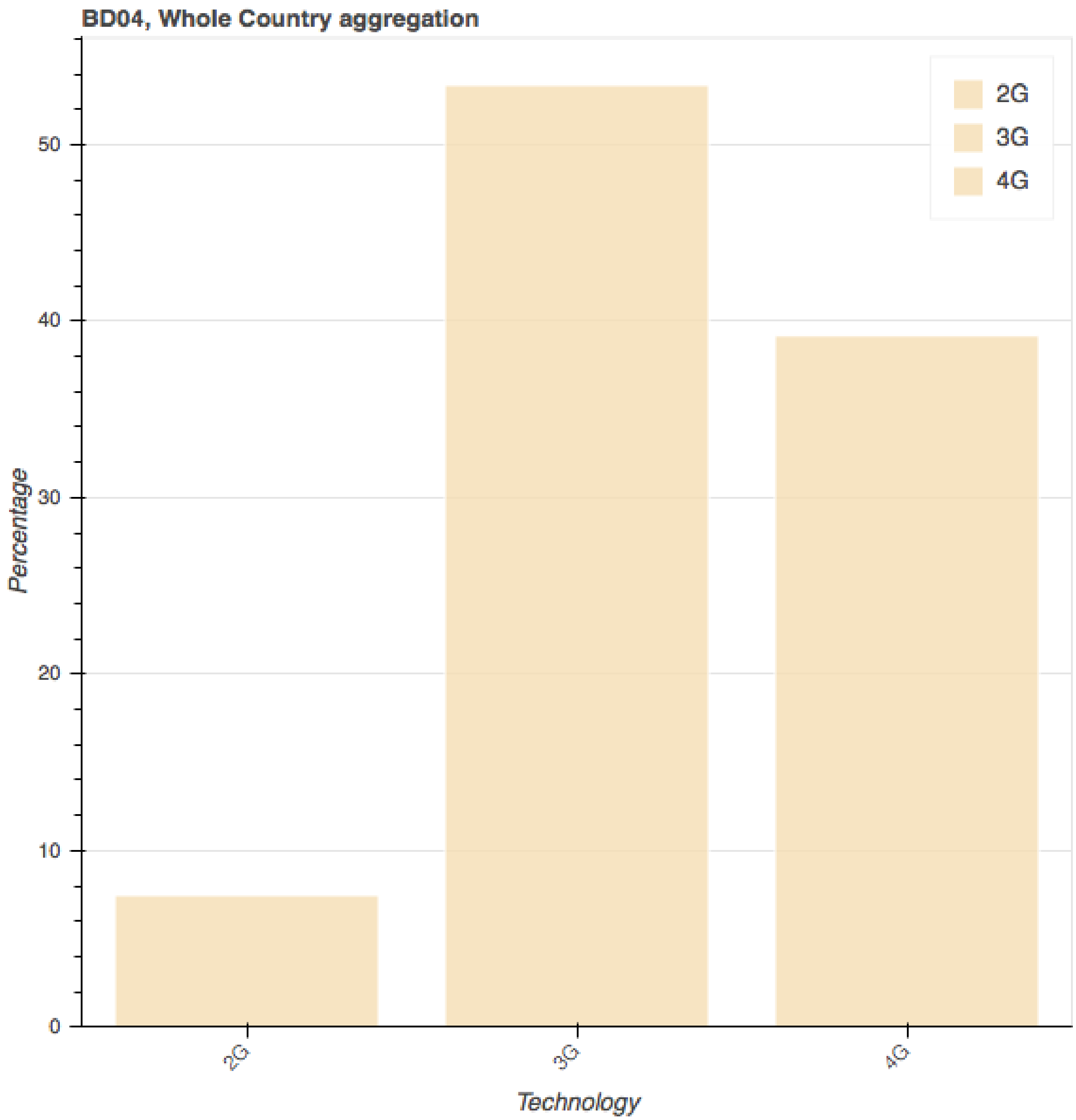
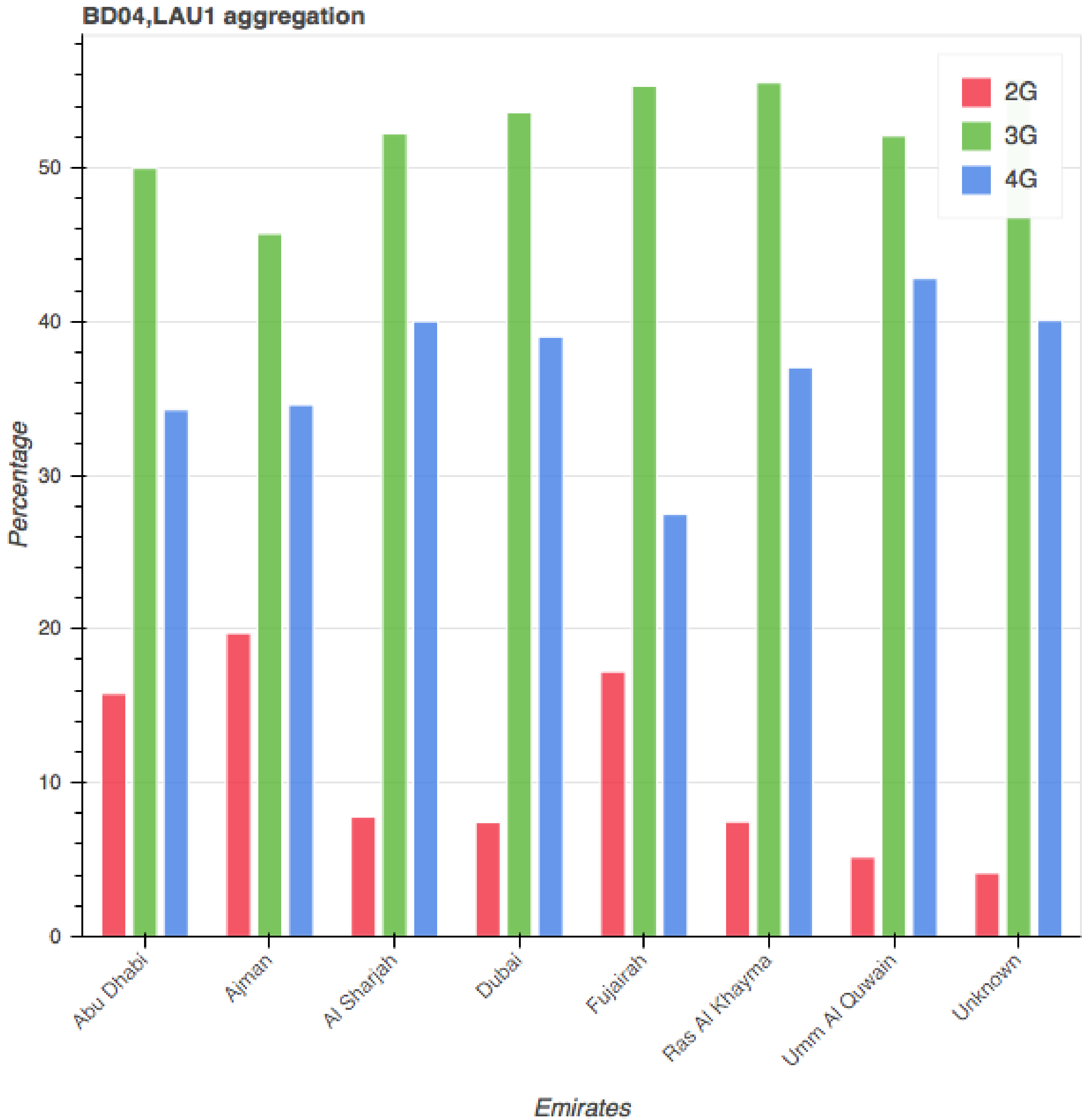


Figure 8: Distribution of mobile Internet access technology according to administrative areas



This indicator is also useful to analyse the mobile Internet access in subregions of the country, thus providing a comparison of the quality of the connectivity across different subregions. In the era of Internet of things and mobile services, it is a useful indicator for business developers, in particular to have a reliable metric to take into consideration when planning new services to provide through the mobile Internet network. Conversely, to the BD03 indicator, here 4G is exploited to provide services for IP-related Internet activities.

Although it is not realistic that 2G network usage for Internet access will diminish in the near future, the urbanized society of the United Arab Emirates shows very well the successful transformation of usage of technologies from 2G to 3G and from 3G to 4G. If such an indicator was collected over a period of time, it should be observed how usage of 2G and 3G technologies are decreasing compared to newer 4G and, probably in the near future, 5G technologies.

The differences of usage by technology among different Emirates show that there are regional differences in the spread and usage of technology that indicate a need for investments in new technologies in such regions. However, the United Arab Emirates could be seen as a good example of shifting technologies with more to achieve, since 4G remains less used than 3G for Internet access.

This indicator does not show the usage by volume and speed (although with the initial raw data, such a possibility can be presented), but it can be assumed that, because of 3G and 4G speed differences, the amount of data transferred via 4G is most likely several times higher that of 2G and 3G.

In addition, as telecommunication service providers cannot geographically allocate some of the antennas in their network to a specific Emirate, the group “Unknown” is also provided in the resulting dataset.

Concerning the decisions that could be made based on regular monitoring of this indicator over time, this could show to both the service providers and the Government the need for investments in technology to equalize the regional differences as well as the country’s level compared to other countries.

5.3. BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

This indicator provides the understanding of how many mobile network subscribers actually have access to and use specific mobile-cellular technology by type of generation (2G, 3G and 4G). Having access is considered a fact of using specific technology over a time period. If a subscriber has only used 2G network antennas for accessing voice, messaging and data service, then this means that subscriber’s access has been limited to 2G technology only and he/she is not able to use higher technology due to technological limitations, service levels or because of contract limitations. The idea of this indicator is to provide insight on the best technology (2G<3G<4G) accessed and used by all unique subscriptions – what is the proportion of the subscribers that has been able to access 2G only (and not above); 3G (and not above); and 4G.

Figure 9: Distribution of subscriptions for each technology type for the entire United Arab Emirates

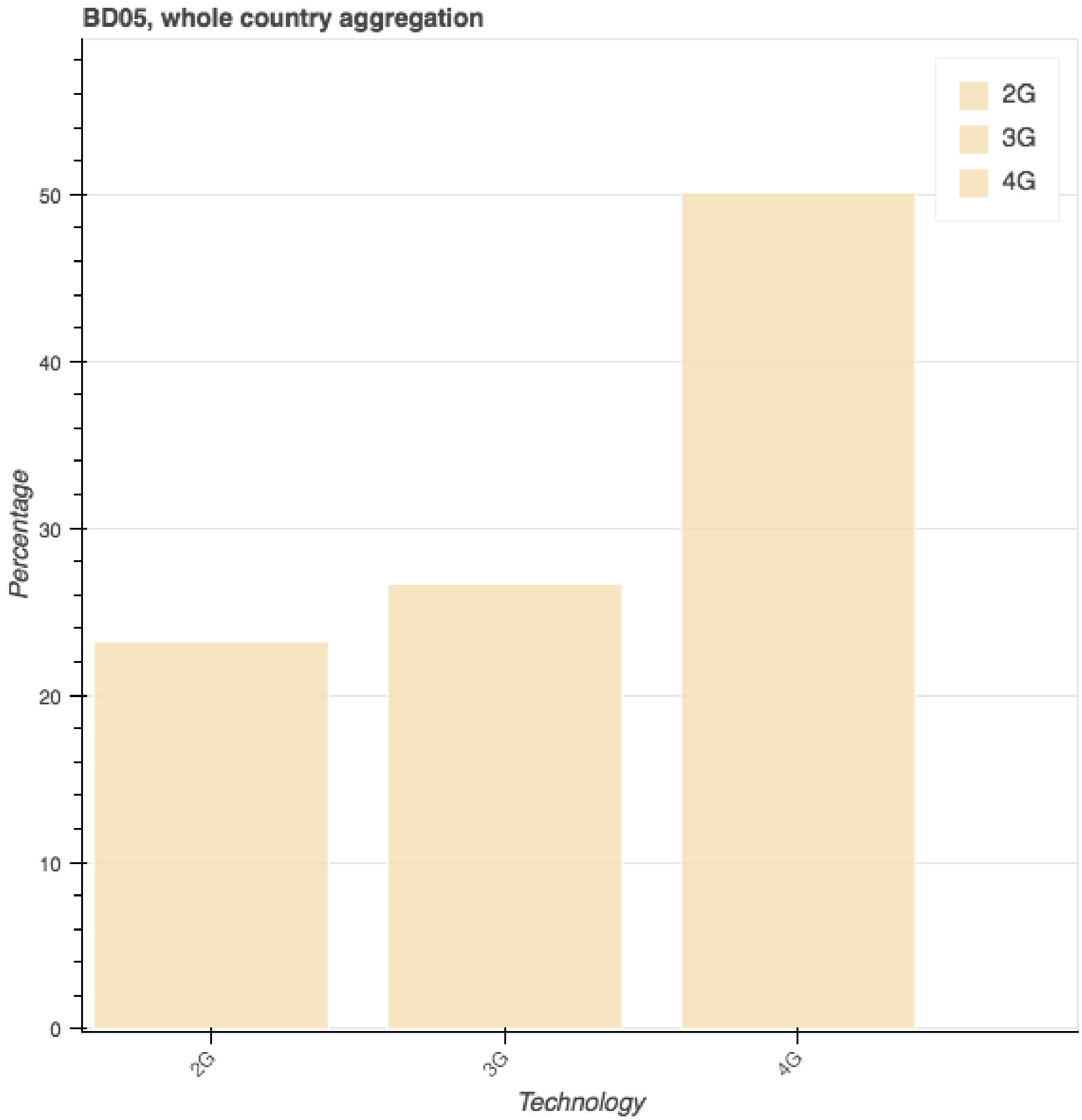
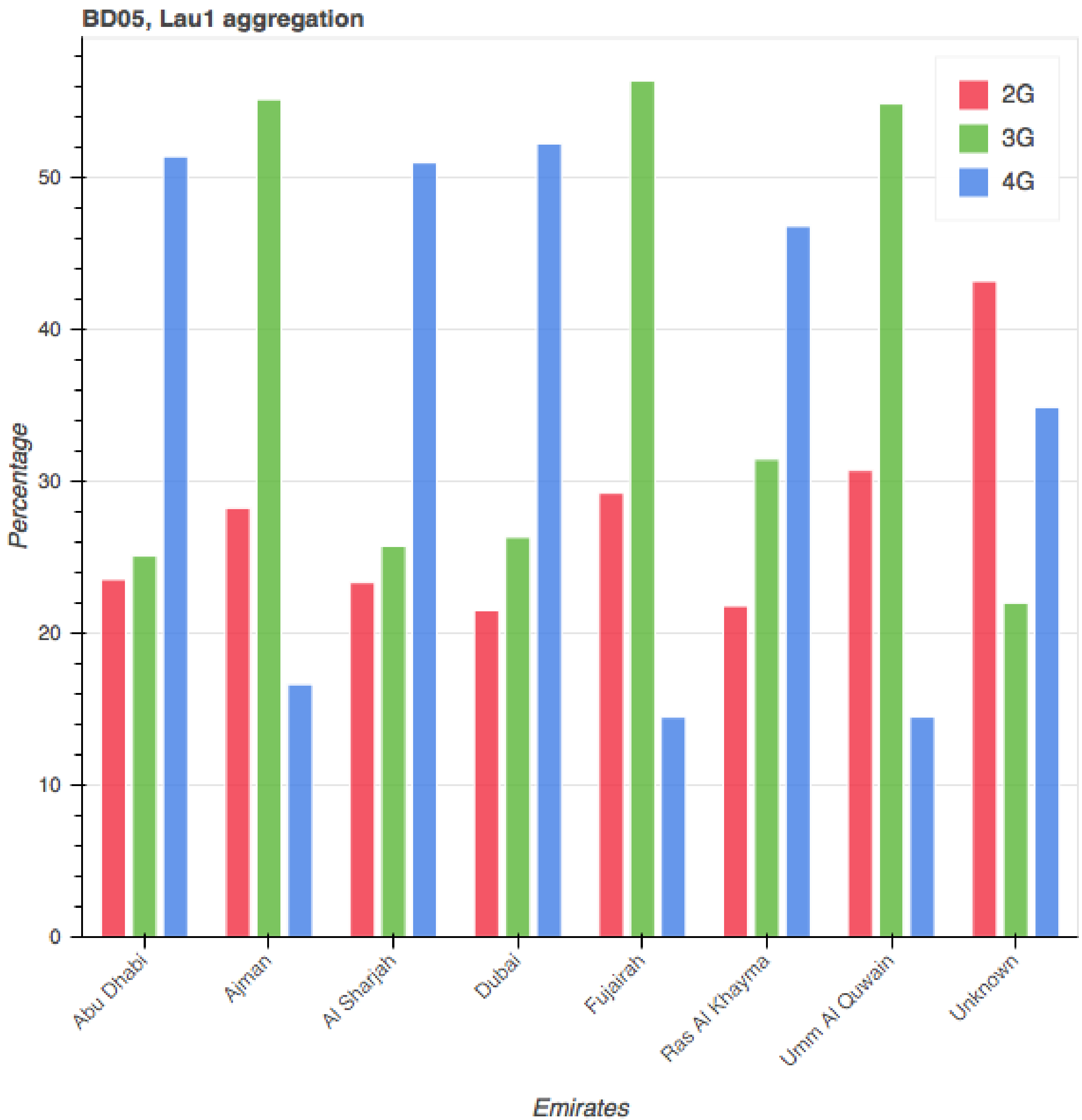


Figure 10: Distribution of subscriptions for each technology type and administrative region



In Figure 10, the distribution of subscriptions for each administrative region is provided. The data show that 4G technology is mostly accessible in Abu Dhabi, Dubai and Al Sharjah, the most densely populated and productive regions of the country. This means service providers have invested heavily in 4G technology in these regions. However, it can also be observed that, in several Emirates, there are many subscribers who do not have access to 4G, and in some cases also a considerable proportion of subscribers who only access 2G networks. Again, there is no ideal threshold or distribution of access, but ideally, 4G access would be 100 per cent with practically no subscribers who can only access 3G and 2G networks.

A practical assessment can be made by comparing this indicator with BD04. For example, although a majority of subscribers in Abu Dhabi can access 4G networks (>50 per cent), most of the Internet usage is still conducted using 3G (50 per cent) and 2G (18 per cent) networks, meaning that access to 4G networks is not permanent, but occasional.

5.4. BD06: ACTIVE MOBILE VOICE AND BROADBAND SUBSCRIPTIONS, BY CONTRACT TYPE

This indicator refers to the number of active mobile subscriptions disaggregated by the type of the contract (prepaid, postpaid; voice and data, data only).

The total number of active mobile subscriptions in the United Arab Emirates was 19,133,341 as of June 2016, according to TRA. This figure is the reported number of subscriptions by service providers to TRA. The number of subscriptions that conducted any activity (calls, messaging and Internet access) during the three-month period (April, May and June 2016) was 22,585,061. There is difference between the active subscription count reported to TRA and in current indicators.

Figure 11: Number of active subscriptions by type for the entire United Arab Emirates

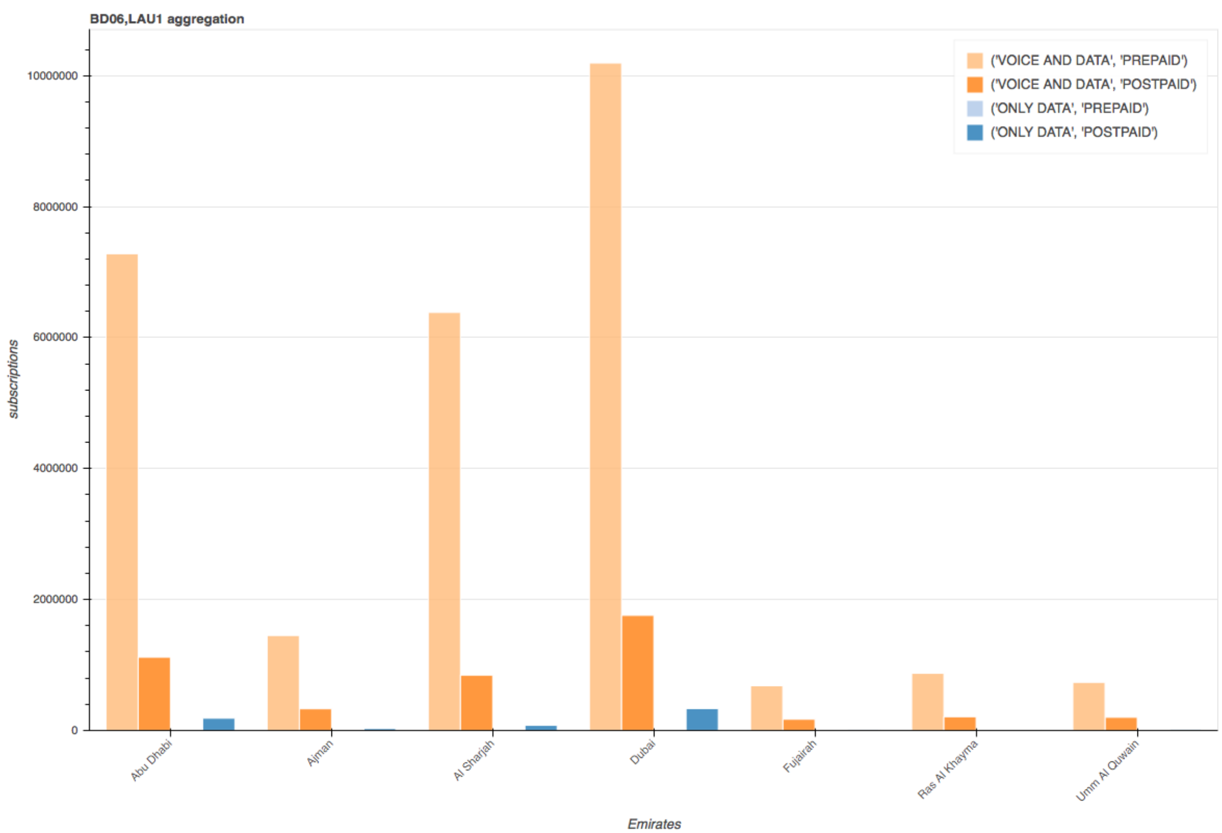
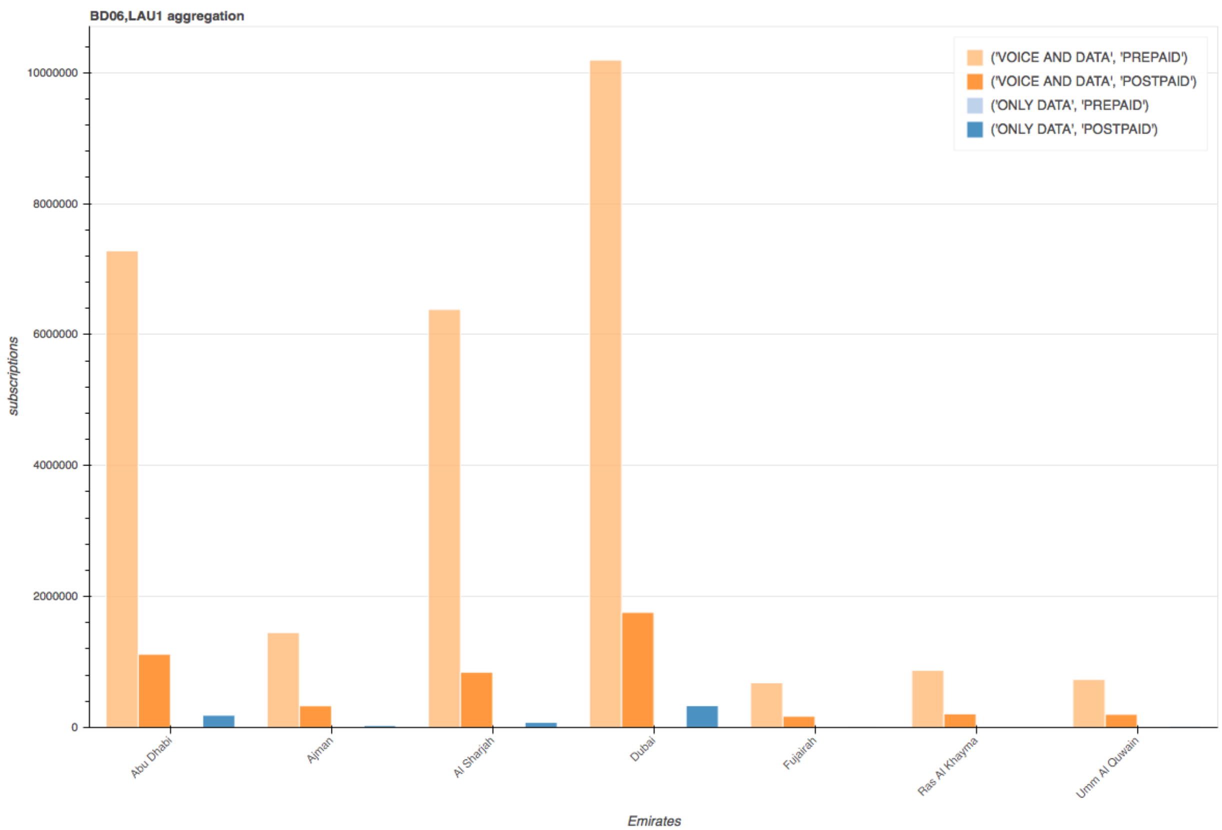


Figure 11 shows the distribution of the number of active subscriptions by contract type. The large majority of contracts are private, postpaid with voice and data service. The second most common case is represented by commercial prepaid voice and data contracts.

In Figure 12, this indicator is disaggregated by administrative regions. Distribution of contract types is similar to the global distribution in Figure 11. As an additional insight, it is possible to note the location of the subscriptions. Dubai is the region where the majority of the events are registered, while Abu Dhabi and Al Sharjah are, respectively, second and third, according to the number of contracts.

Figure 12: Number of contract types (data only, voice and data; prepaid, postpaid) for each administrative region



Comparing the number of active mobile subscriptions that have used the mobile network to the reported number shows how many so-called inactive subscriptions there are in the country/region. In many countries (as in the United Arab Emirates), the penetration rate of the subscriptions is several times higher than the population count, indicating that many people are probably using several mobile devices and subscriptions (e.g. mobile phone, 3G-enabled tablet, etc.). Ideally, this indicator should be the same as the reported number of active mobile-cellular subscriptions (SIM card used at least once in the previous three months for making or receiving calls, carrying out a non-voice activity such as sending or reading an SMS, or accessing the Internet, over the mobile-cellular network), but practically there is a big difference. The distinction by different contract types enables one to see where the missing gaps among the subscriptions are and to target those that are not actually active subscriptions.

5.5. BD08: ACTIVE MOBILE DEVICES

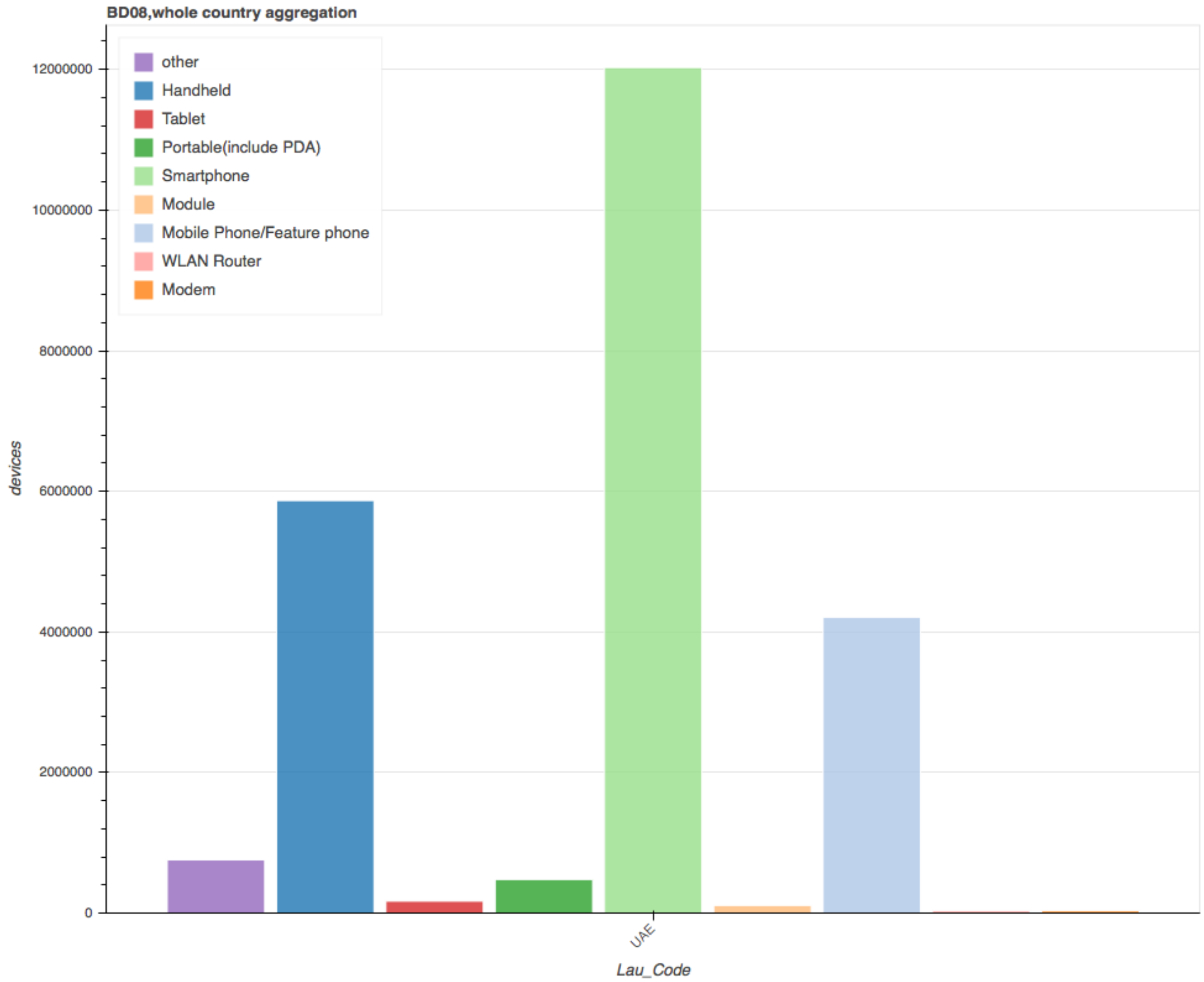
This indicator represents an estimation of the total number of active devices using the mobile network, according to their category. The category of a device is obtained through its International Mobile Station Equipment Identity (IMEI) code and respective Type Allocation Code.

As seen in BD06, there are 22,585,061 active subscriptions and 23,697,919 different devices that have been accessing cellular network service over the three months of pilot period. The total number of different models of devices registered during the observation period is 69,036. This number is calculated based on the unique Type Allocation Code of the devices based on International Mobile Subscriber Identity.

The Type Allocation Code list of different devices allows the distinction of following devices:

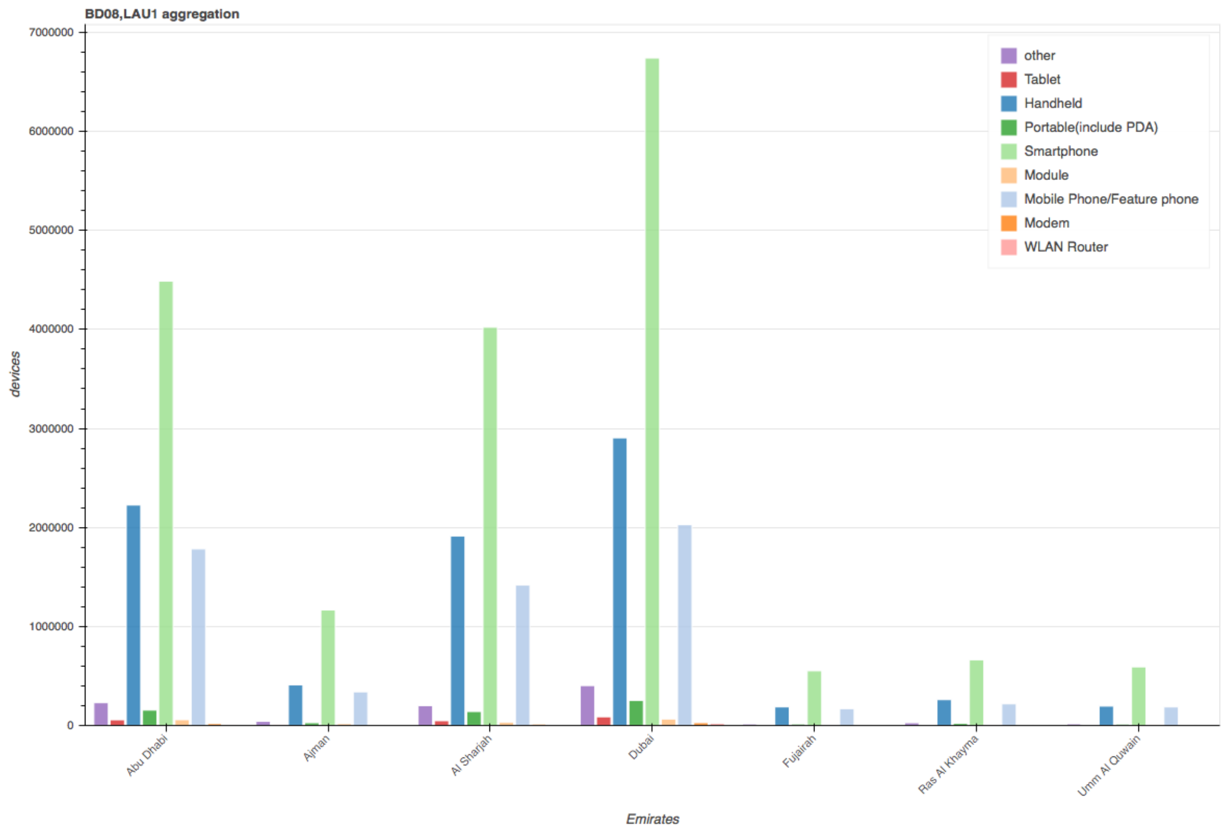
- Dongle;
- Handheld;
- Mobile phone/feature phone;
- Modem;
- Module;
- Portable (including PDA);
- Smartphone;
- Tablet;
- Vehicle;
- Wireless local area network router.

Figure 13: Number of active devices, by type, for the entire United Arab Emirates



In Figures 13 and 14, data for indicator BD08 are reported, aggregated respectively at the LAU0 and LAU1 levels. As expected, smartphones play the main role, being the most used device during the second quarter of 2016. The impact of smartphones is even higher, since handheld devices are the second most active device. Handheld devices are quite similar, in terms of functionalities, to smartphones.

Figure 14: Number of active devices, by type, for each administrative region



The number of active devices compared to the number of active subscriptions shows that there are not so many different devices actually used. By comparing the results with the next indicator (BD09), it seems there are many subscribers who have active subscription and use the device with different SIM cards (the number of devices per subscriptions is 23,697,919/22,585,061 = 1.05). However, the BD09 indicator shows that 14 per cent of subscribers use several devices with their SIM cards. This can be caused by the use of multi-SIM devices, but might require further investigation.

Although it is impossible to provide the exact number of phone users and owners from these data sources, this indicator can be used to estimate such a figure.

5.6. BD09: IMEI CONVERSION RATE

This is a new indicator describing the conversion of mobile devices in the country's mobile networks. Ideally, the IMEI code represents unique devices used in the network. Subscribers change the devices over time, so this indicator should ideally represent the number of subscribers who have changed devices. Because IMEI codes can be changed by using often illegal tools to introduce stolen devices and to avoid IMEI blacklists, the high rate of IMEI conversion can suggest a high level of black market use for stolen devices. The normal or typical conversion rate (the proportion of individuals changing the devices) during a quarter of year should be around 10 per cent (estimation ~5 billion phone users around the world, ~500 million new devices acquired each quarter). If this indicator is much higher, then it suggests the “black market” proportion is high. If this indicator is lower, then customers do not tend to acquire new devices very often.

Figure 15: The percentage of IMEI change during the second quarter of 2016 for each administrative region

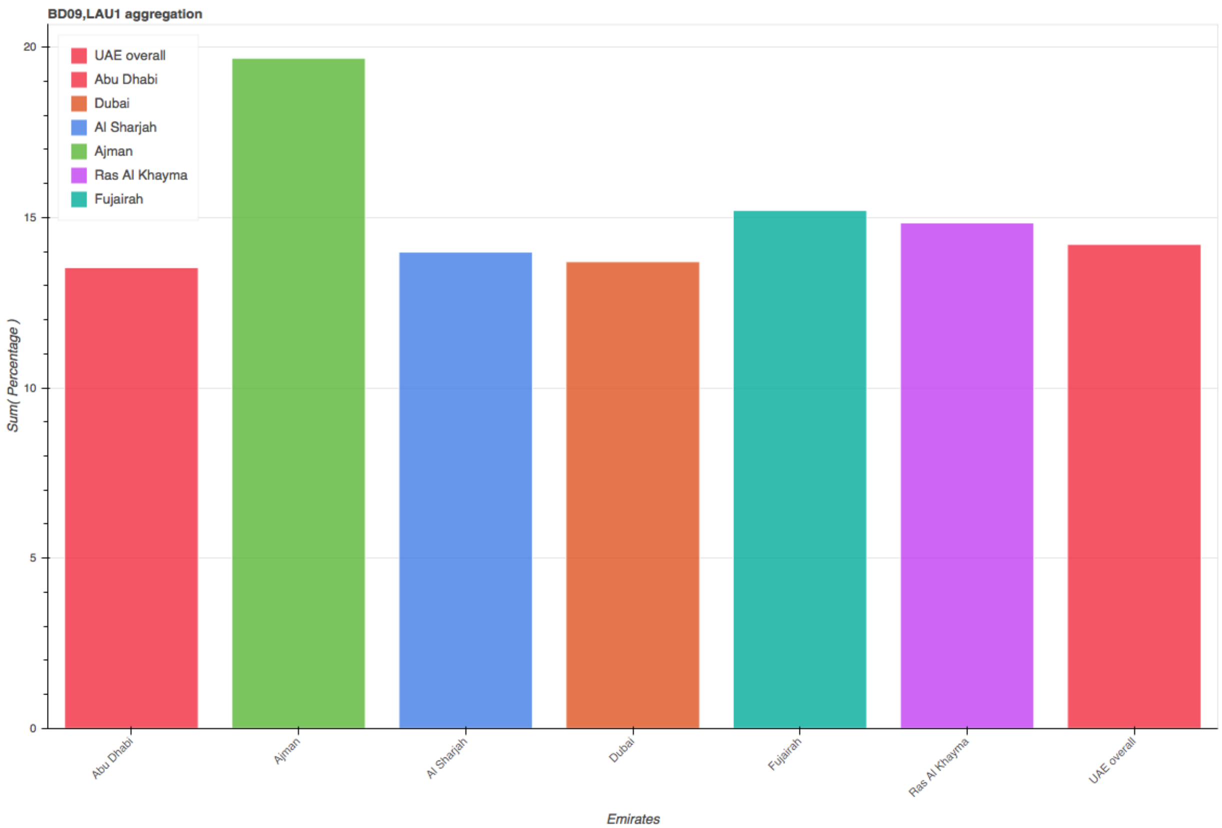


Figure 15 shows the percentage of users changing their IMEI during the second quarter of 2016, around 14 per cent for the entire United Arab Emirates, with a higher value, surprisingly, for Ajman, where it was almost 20 per cent.

This indicator should represent two phenomena:

- If the conversion rate is unreasonably high (although there is no practical guidance as to what is normal and high), then there is a possibility that there are many black market phones in the network, for which IMEI codes are regularly changed to avoid IMEI blacklists.
- If the conversion is low, it can suggest that subscribers are not eager to buy new devices.

For the United Arab Emirates, as mentioned, the overall conversion rate is reasonable (14 per cent), with only one Emirate (Ajman) in which the rate is considerably higher than in other Emirates. Not knowing the local circumstances, it is impossible to make a clear assumption of what is happening in Ajman.

The practical use of this indicator should be investigated and the results should be compared across countries. From the results in the United Arab Emirates, it seems there is a reasonable number of subscribers who change their devices.

5.7. BD11: MOBILE DOMESTIC BROADBAND TRAFFIC, BY CONTRACT TYPE, TECHNOLOGY

Mobile-broadband Internet traffic (within the country) refers to broadband traffic volumes originated within the country from mobile networks. Download and upload traffic should be added up and reported together. Traffic should be measured at the end-user access point. Wholesale and walled-garden traffic should be excluded.

Figure 16: Traffic volume for mobile Internet activity for each administrative region, by technology (volumes are measured in megabytes)

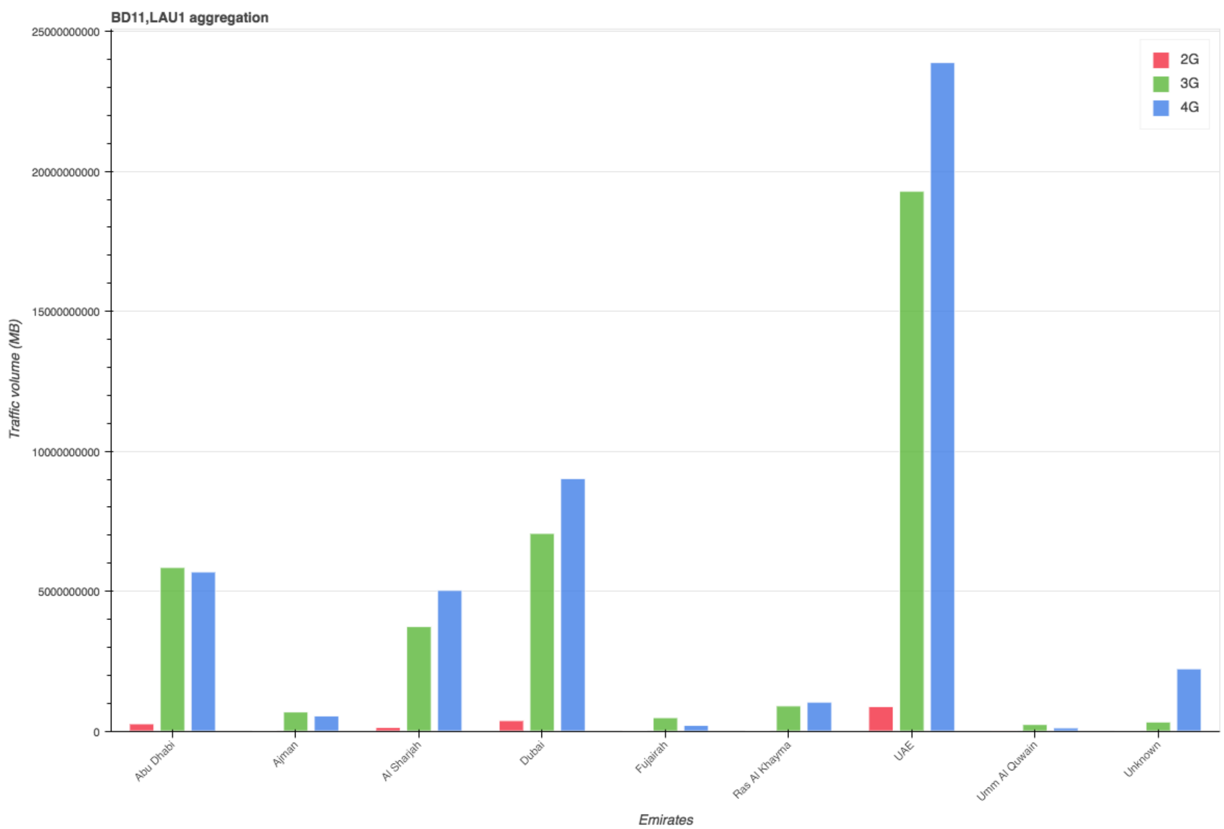


Figure 16 shows traffic volume according to speed and contract type for each Emirate. Again, the Emirate generating the highest volume of traffic is Dubai, both for 3G and 4G. The 3G and 4G breakdowns are comparable in traffic volumes: in some cases, 3G is higher than 4G, but it cannot be foreseen that this difference will be increased in the near future.

5.8. BD12: MOBILE INTERNATIONAL BROADBAND TRAFFIC, BY CONTRACT TYPE

Mobile-broadband Internet traffic (outside the country, roaming out) refers to broadband traffic volumes originated outside the country from 3G networks or other more advanced mobile networks, including 3G upgrades, evolutions or equivalent standards in terms of data transmission speeds. Traffic should be collected and aggregated at the country level for all customers of domestic subscribers' outbound roaming service outside the country and using 3G or more advanced mobile networks.

Download and upload traffic should be added up and reported together. Traffic should be measured at the end-user access point.

Figure 17: Traffic volume for mobile Internet generated by domestic subscribers in the United Arab Emirates and abroad (volumes are measured in terabytes)

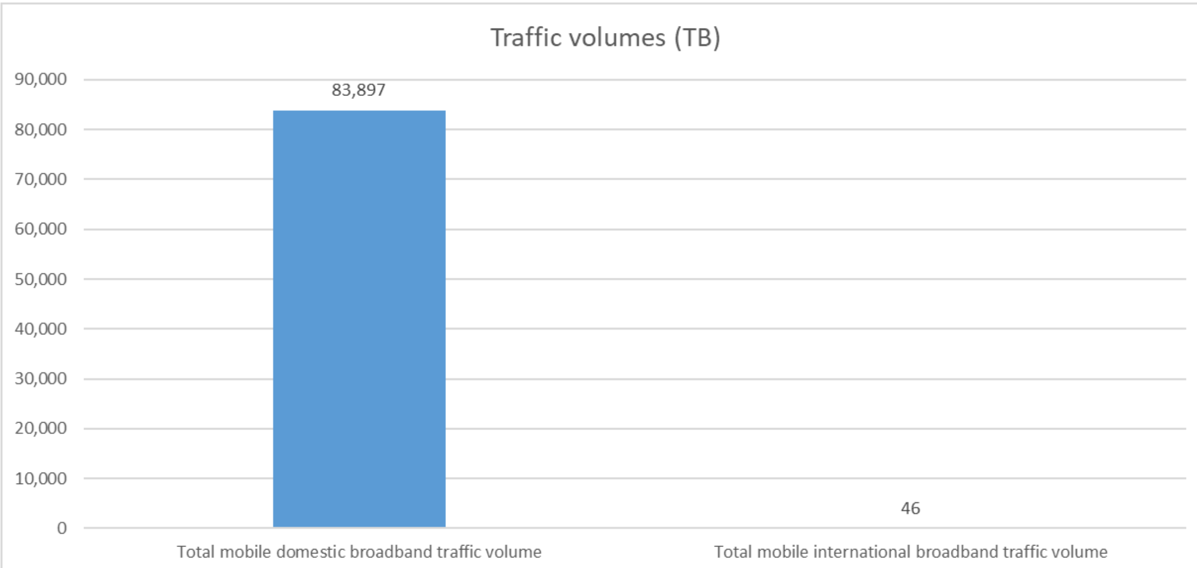


Figure 17 shows the data traffic generated by domestic users while using the roaming service in the United Arab Emirates and abroad. The difference regarding domestic traffic is on the order of 10^4, so domestic traffic is on the order of exabytes, while foreign traffic is in the order of terabytes. It can also suggest that only few subscribers actually access mobile Internet while being abroad, because of the high roaming costs.

5.9. BD14: FIXED BROADBAND SUBSCRIPTIONS, BY TECHNOLOGY

This indicator represents fixed (wired)-broadband subscriptions, aggregated by technology. The indicator can be broken down as follows:

- DSL Internet subscriptions (Copper);
- Fibre-to-the-home/building Internet subscriptions (FTTH);
- Cable modem Internet subscriptions (other);
- Wireless-broadband subscriptions (WiMAX).

Figure 18: Fixed wired broadband subscriptions by technology, for each administrative region

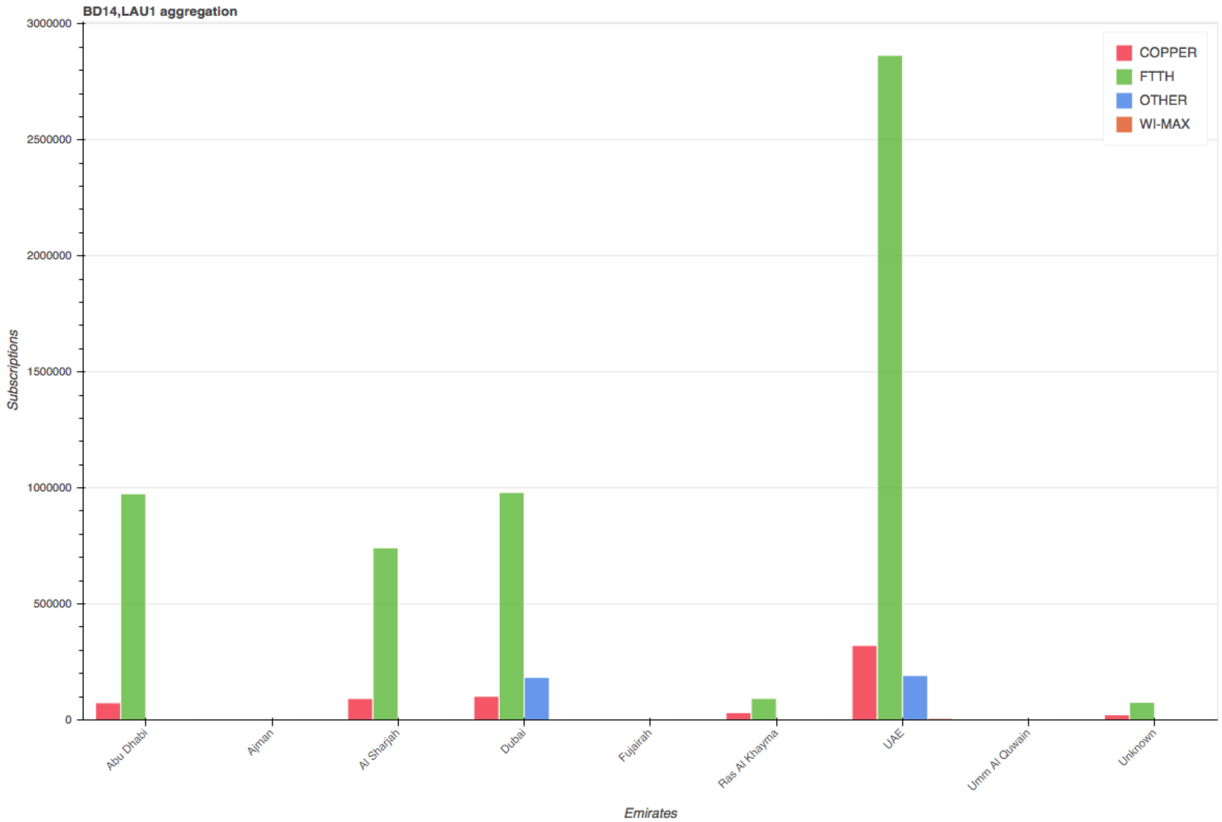


Figure 18 shows the number of broadband subscriptions. Despite the smaller dimension of this Emirate, the number of subscribers with Copper technology from Sharjah is almost the same as in Dubai and Abu Dhabi. In all Emirates, the technology distribution is mostly similar, suggesting that the level of connectivity (in terms of technology used) is very similar in all parts of the United Arab Emirates.

5.10. BD15: FIXED BROADBAND SUBSCRIPTIONS, BY SPEED

This indicator represents fixed (wired)-broadband subscriptions, split by advertised download speed. The indicator can be broken down as follows:

- 256 Kbit/s to less than 2 Mbit/s subscriptions;
- 2 Mbit/s to less than 10 Mbit/s subscriptions;
- 10 Mbit/s to less than 100 Mbit/s subscriptions;
- 100 Mbit/s to less than 1 Gbit/s subscriptions;
- Above 1 Gbit/s subscriptions.

Figure 19: Number of subscriptions according to speed, for each administrative region

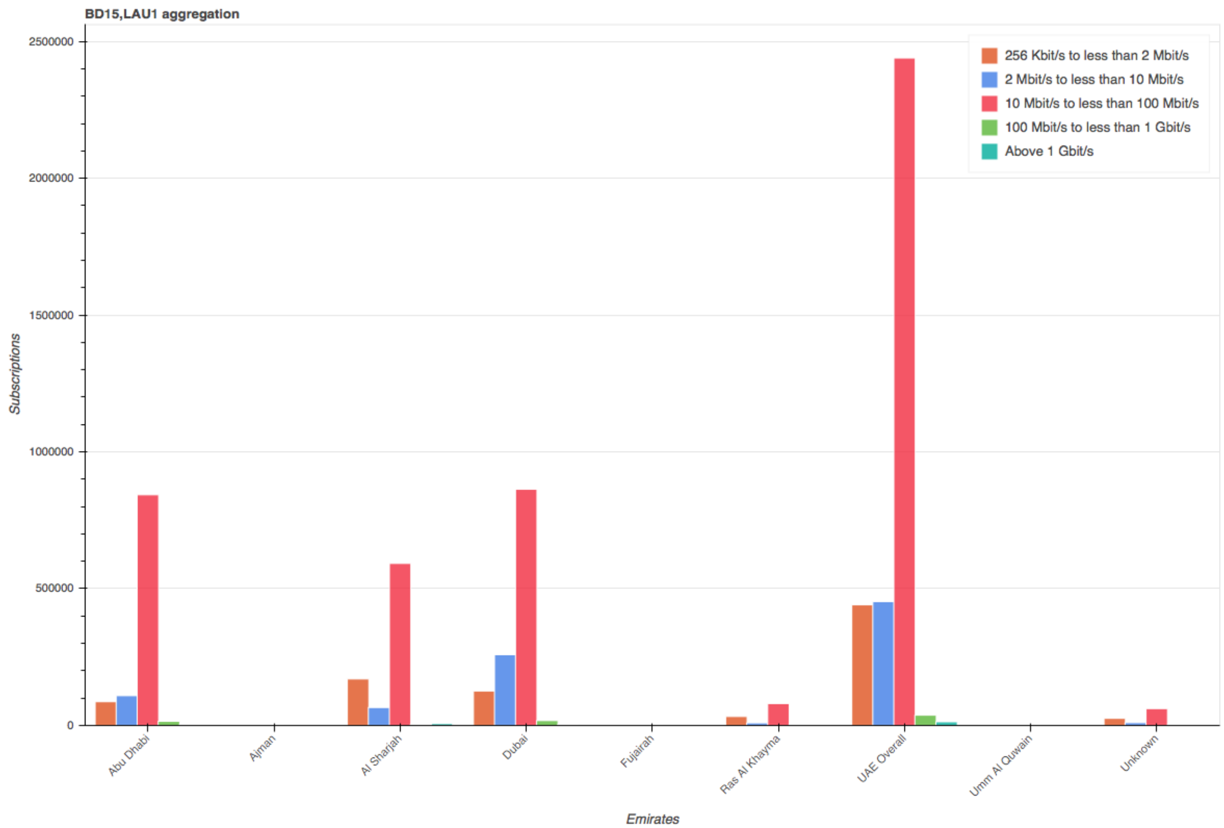


Figure 19 shows the number of subscribers according to speed. Dubai still has the highest number of subscribers, but the number of "low-speed" subscribers is higher with respect to Abu Dhabi. In all Emirates, the speed distribution is mostly similar, suggesting that the level of connectivity (in terms of speeds) is very similar in all parts of the United Arab Emirates.

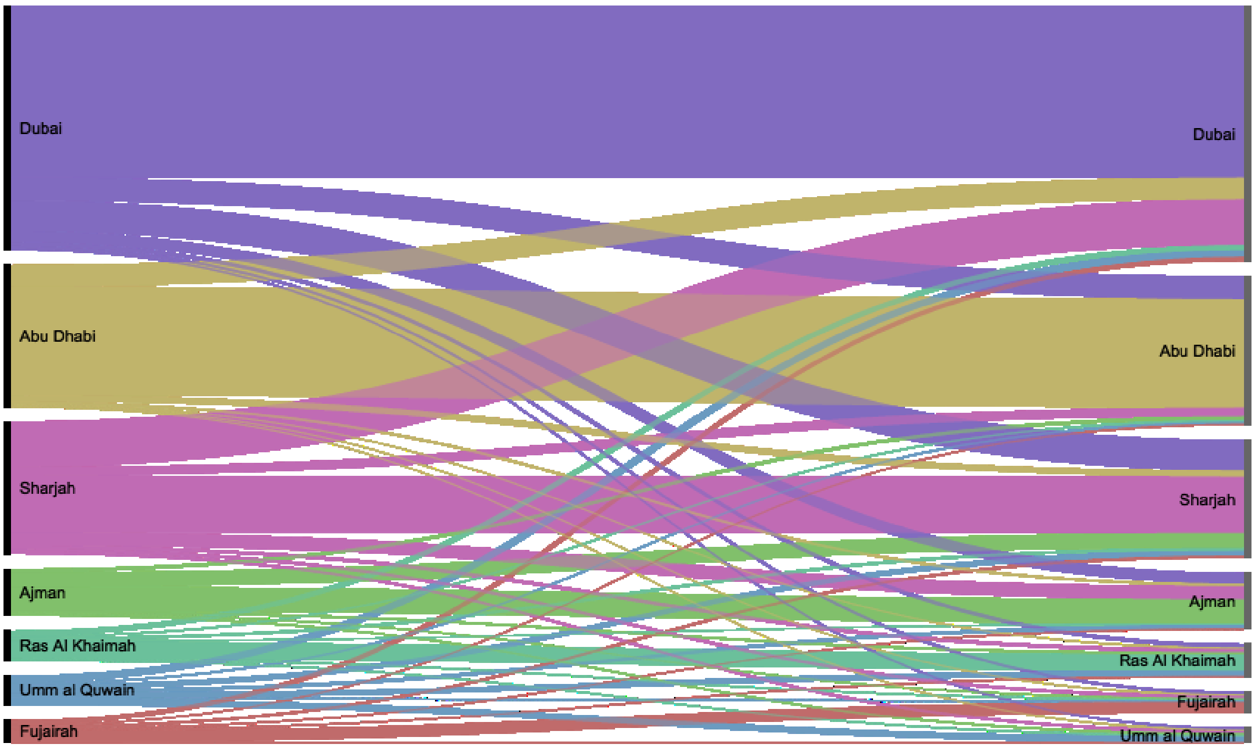
5.11. BD16: ORIGIN/DESTINATION MATRIX

The origin/destination matrix is an indicator widely used in the analysis of human mobility. It is an estimator of the flow between two regions, and it is an estimator for the relationships between such regions. Given N different spatial regions, the origin/destination matrix is an NxN matrix where cell <I,J> represents the number of persons who moved from region I to region J. The coarse granularity of CDR data requires such indicators to be computed on a daily basis as the minimum reliable aggregation. Furthermore, this indicator would require a new type of source that is the most frequent location of the user. In literature, such a feature represents the home location of a user.

This indicator was calculated based on the consecutive events of the subscribers in different Emirates per month. For each subscriber and each month, the "home Emirate" is represented by the most visited Emirate, i.e. the Emirate with the highest number of corresponding CDRs from the specified subscriber. Each CDR belonging to an Emirate different from the home Emirate represents a visit to that Emirate, thus representing the fact that the subscriber had travelled from his or her home Emirate to the next one. This indicator can be further developed and represented as an average, maximum travel counts per day of the week, weeks, national holidays, traffic peak hours, etc., that would be very useful for transportation planning purposes.

Figure 20 shows a normalized origin/destination matrix between each administrative region. The size of the link between two regions indicates the percentage of travellers going from one region to another.

Figure 20: Origin/destination matrix for the United Arab Emirates administrative regions, related to the second quarter of 2016



From the figure, it can be seen that Dubai is the biggest attractor of the country. Most of the people going to Dubai come from the close region of Sharjah, and from Abu Dhabi. Ajman has strong links with Sharjah and Dubai, while Abu Dhabi receives most of the traffic from Dubai. It is worth noting how the size of the flow from Abu Dhabi to Dubai is larger than the flow from Dubai to Abu Dhabi.

Figure 21: Flows of people between Abu Dhabi, Dubai and Sharjah for each month of the second quarter of 2016

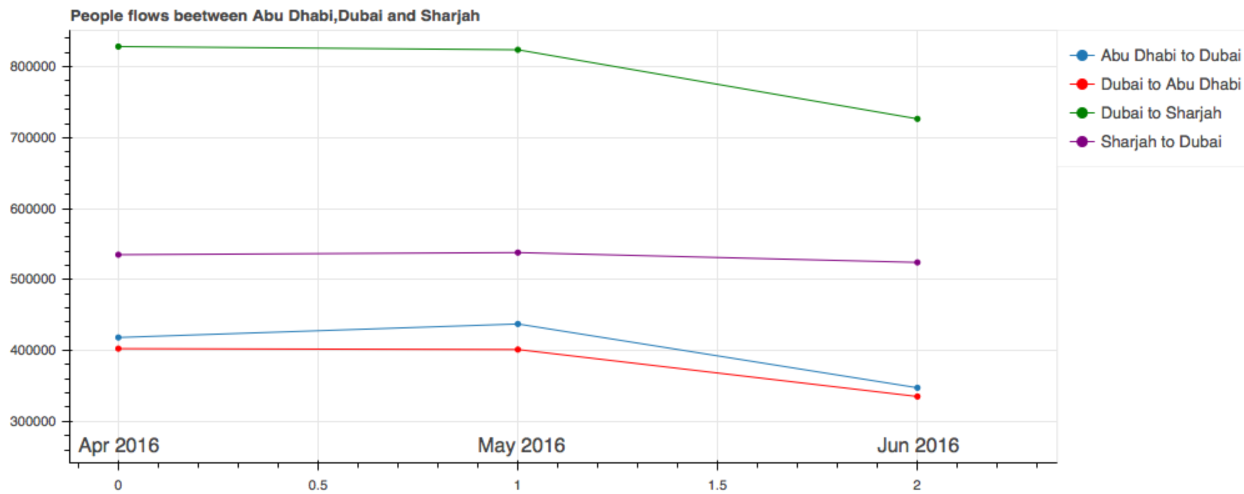


Figure 21 shows the number of people traveling across Dubai, Abu Dhabi and Sharjah. This plots the links between the three biggest regions of the country. The biggest flows are from Dubai to Sharjah and vice versa. This is expected as the city of Dubai is close to Sharjah. The trend that emerges is a clear decrease of all travellers in June, probably due to the period of Ramadan.

Origin/destination matrices can also be used to assess the incoming traffic to the United Arab Emirates from foreign countries. In this case, obviously, CDR data only allow computing the number of users travelling to the United Arab Emirates by taking advantage of detecting roaming CDR. Conversely, roaming CDRs produced by United Arab Emirates citizens in foreign countries are not available.

This indicator represents the value of CDR big data in terms of knowledge: A single, yet huge, dataset is able to provide a complete overview on human mobility across a territory. Such information is useful for several tasks, from mobility planning to tourism analysis and business development.

6. RECOMMENDATIONS

Based on the experience of the pilot project in the United Arab Emirates, the following recommendations can be made:

- Identifying the necessary agreements required for conducting this type of project, or continuous data request procedure, needs to be done well in advance of the project start (memorandum of understandings, NDAs, official request letters, contracts, etc.). It would be very good to have standard legal and administrative procedures in place when the telecommunication regulator and/or national statistics office wants to request indicators based on big data in the future.
- The proposed standards in naming and coding conventions should be followed by all telecommunication service providers. Cleaning and reformatting the data by the receiving party can take a lot of time, and ideally this procedure has to be automated, so agreeing on standards of the data format and data exchange is important.
- The reference data required for the data processing should be identified, obtained and formatted for the processing. The reference data include the geographical data about administrative subdivisions (multiple levels), population data (grid-based or administrative units), any existing indicators that are to be combined with big data indicators (e.g. BD13), etc. The lack of two pieces of reference data, or using different types of reference data, may affect the quality of the resulting indicators (e.g. if different telecommunication service providers use different administrative levels, units or coding, the data cannot be combined).
- Validation methods of the steps of processing the data (extraction, processing and results) should be agreed between all stakeholders. How to validate the data extraction and processing of providers (and processors if they are different parties) is a rather complicated question. Here the validation of extracted data (distributions, descriptive statistics, etc.) and confirmation of exact algorithms used for processing can be the option, but how to practically conduct that can be an issue (given that these might be sensitive business secrets of the telecommunication service providers). The validation of results is often problematic if there is limited information on the technical algorithms used for calculating these indicators. It is important that national statistics offices take part in the validation of the indicators.
- It is important to understand that, in practice, the development of the methodology is a multiple iteration of preparing the methodology, creating algorithms, producing the results, assessing the results, adjusting the methodology, etc. This can be a time-consuming process.
- In this pilot project, the project coordinator and the data scientist were only able to visit the pilot country once each. This type of project would require multiple visits with intermediate calculations of the indicators, assessing the results and adjusting the methodology and algorithms.
- Agreeing on the source data, methodology, specific algorithms, business model and validation procedures takes a lot of time, so it is reasonable to plan an intensive preparation period, including several meetings between stakeholder businesses, legal and technical teams. During this project, the time for this preparation was very limited, but this can be considered a learning opportunity – this is a pilot project designed to identify such issues.

- Agreeing on the scope of the publication of the resulting indicators in terms of confidentiality of the telecommunication service providers is important, especially if the number of providers is small, e.g. to determine the threshold of confidentiality of the published data and how to ensure the safety of the business secrets of the telecommunication service providers. It is important in terms of public image that local project organizers (focal points) position themselves as the guardians of the confidentiality of their informants (telecommunication service providers). Thus, it is important to exercise great responsibility in the management of the risk of disclosure confidential information.
- It is also important to have a strong coordinating mechanism to determine the roles and responsibilities for all stakeholders involved in the project, including ensuring that the responsible stakeholders have the appropriate resources, skills and time allocated for conducting the necessary tasks. Commitment of all stakeholders in an agreement should be signed to ensure the participation of the stakeholders in the project.
- It is recommended to obtain from the telecommunication service providers an initial assessment of the resources required and their availability to process the data within the agreed time-frame.
- This pilot project provided some of the indicators that can be calculated from big data in ICT, but more discussion on the usefulness, practical applicability and further development of those indicators is required. Based on the resulting indicators, it is expected that, in each country, some of the indicators can be important and used for national purposes, and some indicators are valuable for international comparability.
- The value of the big data does not simply lie in the production of indicators with more breakdowns; it also provides the possibility to employ different data mining, machine learning, deep package inspections, and other mathematical and statistical methods that were not practiced in this pilot project due to limited time and resources. These more sophisticated methods can provide additional insights for the ICT industry, but using such methods requires more intrusion in the telecommunication service providers' databases and is therefore limited to external parties such as ITU. For example, analyses of the individual's behaviour changes in the consumption of ICT could provide insights into the trends and foresights (prediction) of the use of ICT – not simply by looking at the overall trends of the usage, but by the clustering and classification of the users based on their individual behaviour and activity.
- It is beneficial to involve other potential stakeholders from the public and private sectors, either in the planning process or for presenting the results of the indicators. As seen in the United Arab Emirates, additional indicators that can be related to ICT can also have great value for non-ICT sectors (e.g. origin/destination matrices). Domains such as tourism, transportation, urban and regional planning, social sciences, etc., can largely benefit from this data source. Including universities in the projects can also boost the scientific research potential in a variety of domains and provide valuable feedback for the telecommunication service providers as well as public organizations.
- The promotion of projects including big data is recommended for good publicity and explanation to the public how use of big data can benefit the society at large.
- It is also important to set the objectives of the development in ICT sectors that could be measured with such indicators. For example, the objective could be that, in the next two years,

the number of subscribers with access to 4G and higher technology would be 75 per cent (BD05) for the whole country, and in no administrative subdivisions this number should be below 50 per cent, while the difference between urban and rural areas should not be more than 25 per cent (e.g. urban 80 per cent, rural >65 per cent). The specific target objectives are obviously country-specific and can be agreed between local stakeholders.

7. CONCLUSIONS

In the pilot project, a number of indicators related to ICT were proposed and calculated. There are two main possibilities to assess the value of these indicators:

- On the national level – if the indicators represent new insight into ICT situation;
- On the international level – to compare the ICT levels of different countries.

As the country reports are generated after the results have been received, the comparison between the countries is presented in the Project Final Report. Only then can the comparison between the countries and the international value of the indicators be assessed. The international comparability can also provide additional insight on the value for the indicators within each country.

One of the objectives of the project was to test the process of generating new indicators based on big data, and identify the challenges and possibilities. This information is also valuable as the challenges are often similar in different countries and the results serve as a good insight for other countries that wish to conduct such experiments.

In general, the challenges can be divided into:

- Administrative and legal;
- Technical and methodological.

In the United Arab Emirates, the administrative and legal challenges were successfully dealt with because of the strong commitment of FCSA, TRA and both telecommunication service providers. The technical aspects were not difficult to resolve, mainly because of the readiness and experience of both telecommunication service providers, who deal with big data on a daily basis. However, it must be noted that time allocated and planning could be improved relating to preparation, discussion on the methodology, specific algorithms and issues with standardization of data and indicators from both of the telecommunication service providers.

The methodology provided for this project is a starting point; however, it requires further improvement and standardization. It is crucial to compare the results from other pilot countries to make the overall assessment and understand individual indicators' value.

The resulting indicators in the United Arab Emirates show a rather advanced stage of the ICT – the usage of 3G and 4G is higher than 2G (BD03 and BD04) and it can be seen that half of the subscribers have access to and use 4G services. The increasing trend of moving towards newer technologies is expected – ideally, all subscribers should have access to 4G (100 per cent in BD05), but in reality this is hardly achievable. As there was no information available on the distribution of the population among urban and rural areas, the difference between usage and access of the technologies in urban and rural areas cannot be measured. However, as the majority of the population in the United Arab Emirates lives in the urban areas, it can be considered that all of the subscribers involved are in fact living in the urban areas.

The number of subscribers who have been using voice or broadband services is 22.5 million in the United Arab Emirates, there have been 23.7 million different devices and 70,000 different device models. As the IMEI conversion rate (proportion of subscribers who changed mobile devices during period) for the United Arab Emirates in total is 14 per cent, it shows that, for some reason, there are subscribers who use their SIM card in several devices.

The traffic volumes generated by 3G and 4G technologies is comparable at the moment (2G is very low compared to the newer technologies), but this difference is expected to increase. Travelling subscribers from the United Arab Emirates do not use the Internet much while abroad. The difference between domestic mobile traffic and traffic abroad is 2 000-fold.

Additionally, the stakeholders in the United Arab Emirates were able to generate mobility origin destination matrices for a number of trips undertaken by the subscribers between different Emirates. However, this is not an ICT indicator, as the practice also shows the usability of such data sources in other domains, such as transportation planning. The fact that service providers proposed and calculated such additional indicators shows the willingness and ability of the stakeholders in the United Arab Emirates to be innovative and discover new opportunities outside the ICT domain.

The results of this pilot project show the advanced level of ICTs in the United Arab Emirates. The unavailability of two pieces of reference data suggests possible improvements; however, these are issues not directly related to ICT.