BIG DATA FOR MEASURING THE INFORMATION SOCIETY: COUNTRY REPORT – GEORGIA

CONTENTS

Abbreviations

CDR	Call Detail Record
GNCC	Georgian National Communication Commission
ICT	Information and Communication Technology
IMEI	International Mobile Equipment Identity
IPDR	Internet Protocol Detail Record
ISP	Internet Service Provider
LAU	Local Administrative Unit
LTE	Long-Term Evolution
MNO	Mobile Network Operator
SMS	Short Message Service

Tables, figures

Tables

Figures

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 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 Georgia. 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, and 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, analyzed 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 pilots and make recommendations on how countries and ITU could use the results of the pilot studies for their ICT data collection.

1.2. PILOT COUNTRY CONTEXT

Georgia is ranked 74th, with an ICT Development Index of 5.79 in 2017. Growth of the country's telecommunication market is in line with the slow growth of its economy in general. However, relative to other sectors in Georgia, the telecommunication market remains among the fastest growing and now represents approximately 5%-7% of gross domestic product. The telecommunication market is also undergoing a shift away from traditional voice and messaging services to alternative over-the-top services. Fixed telephony revenues have fallen significantly, accompanied by a slight decrease in mobile retail revenues. On the other hand, other services – such as Internet, broadcasting transit – have experienced increased revenues.

The penetration of mobile telephone subscribers was 148 SIM cards per 100 inhabitants in end of 2016. There are three GSM mobile-cellular service providers – Beeline, Geocell and Magticom – in Georgia. There are more than 10 major ISP providers, including the mobile service providers. Long-Term Evolution (LTE) service has been successfully introduced in Georgia, and extension of the spectrum for 3G and LTE has improved the availability and quality of service over 2016 and 2017. All MNOs are investing to expand the reach and capabilities of their LTE infrastructures to areas outside of Tbilisi.

Since the Georgian National Communication Commission (GNCC) is currently actively collecting and processing big data from telecommunication providers in order to monitor market developments, participation in the ITU pilot project was a logical next step. Although the data collected in this pilot project is larger and more detailed compared to what GNCC normally collects, a similar administrative approach and technical setup were used.

1.3. PILOT PROJECT TIMELINE

The pilot project in Georgia was implemented from May 2016 to October 2017 (Table 1).

Table 1: Project timeline

Activity	Time
Initial communication, stakeholder involvement and planning the project activities	May - July 2016
Kick-off meeting of the project with all stakeholders involved in Georgia	11 August 2016
Preparations for processing the data by data providers, assessing the methodology for calculating the indicators and planning/allocating the resources for the tasks, resolving legal aspects	September 2016 - April 2017
ITU data scientist's mission to Georgia	3 - 14 April 2017
Calculation of the indicators (after the visit of the data scientist)	April - September 2017

Assessing the resulting indicators; acquiring some corrections from data providers; initiating the country report; getting feedback, comments and assessment from local stakeholders in the report	August - October 2017

The timeline of the project has been extended several times due to challenges caused by administrative, legal and data-related issues. The initial plan of finishing the project in 2016 was postponed until the end of October 2017.

2. STAKEHOLDERS IN THE PROJECT

The following local stakeholders were involved in the implementation of the pilot project:

- GNCC, the telecommunication regulator, was the country focal point responsible for coordinating the activities related to the project in Georgia.
- National Statistics Office of Georgia was responsible for providing administrative reference data.
- MNOs: Magticom, Geocell and Mobitel.

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

3.1. LEGAL DOCUMENTS AND CHALLENGES

During the project, the following administrative and legal steps were taken for GNCC to obtain the data from the data providers:

- GNCC made an official request to telecommunication service providers to obtain the data in necessary format and for a specific period.
- GNCC organized several meetings with the service providers to discuss the availability of the data and approve the format in which the data would be presented.
- Data protection issues were discussed with and approved by the Personal Data Protection office, and relevant steps were taken to make the data anonymous; also, specific steps defined by the office for storage and access of the data were approved and followed strictly.

3.2. RESOURCES REQUIRED FOR THE PROCESSING

As the data used in the project were collected directly from the telecommunication service providers and processed using the infrastructure of GNCC, the following efforts and resources were used in the project:

- Data were stored in the server, which required 5TB of storage capacity.
- Data were stored using two databases: MSSQL and PostGIS. The tools used to process the data were MSSQL and PostgreSQL database managers and QGIS application.
- The team involved in the processing of Tier I data into Tier II data included two MSSQL and PostgreSQL database programmers, one IT specialist and a manager.

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 antenna 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 also are used to assess the coverage area over the land and population.
- The MNOs' network usage data log records as CDRs and IPDRs: These records represent domestic, inbound roaming and outbound roaming data. These data were combined with antenna coverage data and include all records individually for incoming and outgoing voice CDRs, incoming and outgoing messaging CDRs (Short Message Service (SMS) and Multimedia Messaging Service), and Internet traffic detail records (IPDRs). Each event (usage log) has a separate individual record (e.g. one outgoing call is one record with attributes). Concerning the IPDR, 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.
- ISPs' fixed broadband subscribers' Internet access statistical information (Source – GNCC electronic statistical database).

CDR is a well-known and standard format for recording subscriber activity (i.e. call, SMS and Internet data usage session). Each record contains attributes such as the time stamp of the beginning of the event, the antenna the device was connected to, the duration of the activity, and some other features, tracked for operational and/or billing purposes.

Table 2: Example of raw data from MNO

ID	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	[illegible]	1	2	2	5	1462215001	736	1	209	[illegible]	
90000004	[illegible]	1	2	2	1	1462419902	1400	1	209	[illegible]	
90000004	[illegible]	1	2	2	6	1462651811	1630	1	209	[illegible]	[illegible]
90000004	[illegible]	1	2	2	2	1462654988	1457	1	209	[illegible]	
90000004	[illegible]	1	2	2	4	1462758074	336	1	209	[illegible]	
90000004	[illegible]	1	2	2	5	1462947150	457	3	209		
90000004	[illegible]	1	2	2	6	1463248854	1093	3	209		[illegible]
90000005	[illegible]	1	2	3	6	1461623716	1077	1	209	[illegible]	[illegible]
90000005	[illegible]	1	2	3	6	1461778309	192	1	209	[illegible]	[illegible]
90000005	[illegible]	1	2	3	6	1462036940	1723	1	209	[illegible]	[illegible]
90000005	[illegible]	1	2	3	6	1462340914	1101	1	209	[illegible]	[illegible]
90000005	[illegible]	1	2	3	6	1462445085	1932	1	209	[illegible]	[illegible]

Table 3: 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]

Three mobile network operators provided CDRs and IPDRs for the pilot project: Magticom (40% subscriber share), Geocell (35% subscriber share) and Mobitel (25% subscriber share). However, the data provided by Beeline came late and could not be properly validated, and were therefore excluded from the calculation of the indicators for the pilot project. The resulting indicators were calculated using data representing more than 75% of subscribers in Georgia. The data covered the fourth quarter of 2016 (October, November and December).

The initial raw data provided by both operators combined had a total of 50 billion rows and 2.7 terabytes.

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 Georgia subdivisions (In Georgia, there are three levels of administrative units: regions (LAU1), municipalities (LAU2) and communities/towns (LAU3)) (source - National Agency of Public Registry of Ministry of Justice).
- A one-square kilometer grid of population. These reference data was used during calculate BD02 (source - National Statistics Office of Georgia).
- Monthly incoming international travelers' statistics by country (source - Georgian National Tourism Administration).
- GSMA Device Database (GSM Association).

4.3. ACCESS TO THE DATA

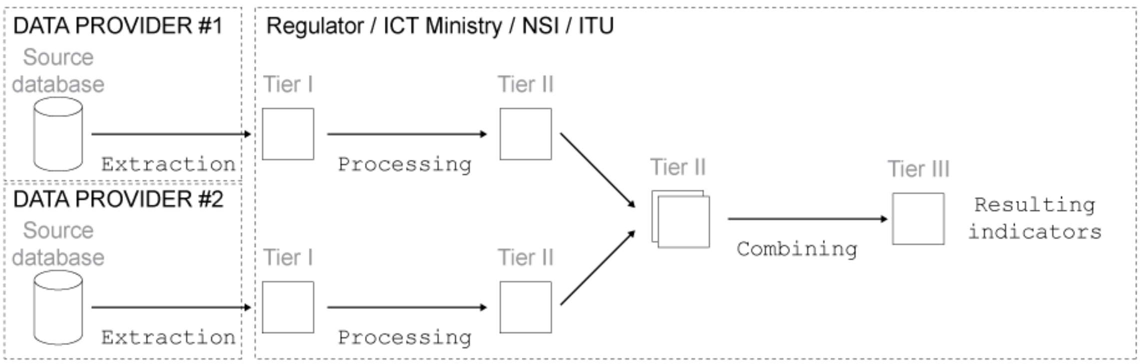
Due to privacy protection reasons, and because initial raw data from the telecommunication service providers can be considered to include confidential business information, there are three tiers or phases of the data processing:

- Tier I – Initial, raw, not aggregated data extracted by the telecommunication service providers from their databases and registries, which is the basis for calculation and which might include private and confidential business information. This can also be referred as micro data.
- 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 macro data.
- Tier III – These Tier II aggregated indicators from different telecommunication service providers are merged and 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 Georgia, the process of extraction and calculation of Tier II indicators was done on the premises of GNCC with Tier I micro data (raw data individual records) provided only to GNCC by telecommunication companies (figure 1).

Telecommunication service providers extracted the initial raw data from their data warehouses, they did anonymize the data, using an anonymized methodology developed within the project, relying on big data processing technologies, providing them to GNCC, which processed the data according to the methodology and aggregated to Tier II and Tier III data, which were stored as CSV files. The CSV files were delivered to the data scientist. Tier I data were stored in GNCC's SQL server, but the data scientist did not have direct access to it. GNCC provided Tier II aggregations to the data scientist and this report is prepared by GNCC in cooperation with data scientist.

Figure 1: Data acquisition steps for Georgia



4.4. PROCESSING THE DATA

Tier I data have been ingested and processed into GNCC's dedicated servers, running an MSSQL and PostgreSQL database servers. All the resulting Tier II indicators available have been deployed to the data scientist through standard CSV files. The computation of Tier III indicators, together with all the aggregations and visualizations provided in this report, has been performed exploiting technologies such as QGIS visualization tools and Bokeh visualization framework.

Tier I data were delivered to GNCC during the data scientist's visit and not before. Because of this, GNCC's IT department had to work very fast during the data scientist's visit in order to import, clean and review the data. Issues regarding coverage of the antennas and consequently, the geographical reference of Tier I data, needed to be checked and confirmed, which time-consuming assignment was. After discovering inconsistencies with the coverage data, GNCC requested telecommunication data providers to fix them and provide back correct Tier I raw data. The ingestion process for new raw data took more than two weeks. The process of checking Tier II data, which was supposed to take place during the data scientist's visit, has been performed remotely, making the entire process slower. Several issues came up, and thus Tier II data were made available to the data scientist three months after the visit.

Tier II data assessing and validation by the data scientist exposed several calculation issues for most of the indicators, which were successively fixed and recalculated by GNCC. For example, the number of subscribers was not initially calculated based on the unique subscriber IDs. This led to an incorrect number of subscriptions in different indicators. Tier I raw data were not accessible by the data scientist, thus making it difficult to identify and understand the issue at hand. This also made the validation and recalculation process slower, every issue pointed out by the data scientist remotely, based on Tier II data, was afterwards assessed by GNCC, until a fix was provided and the indicator recalculated.

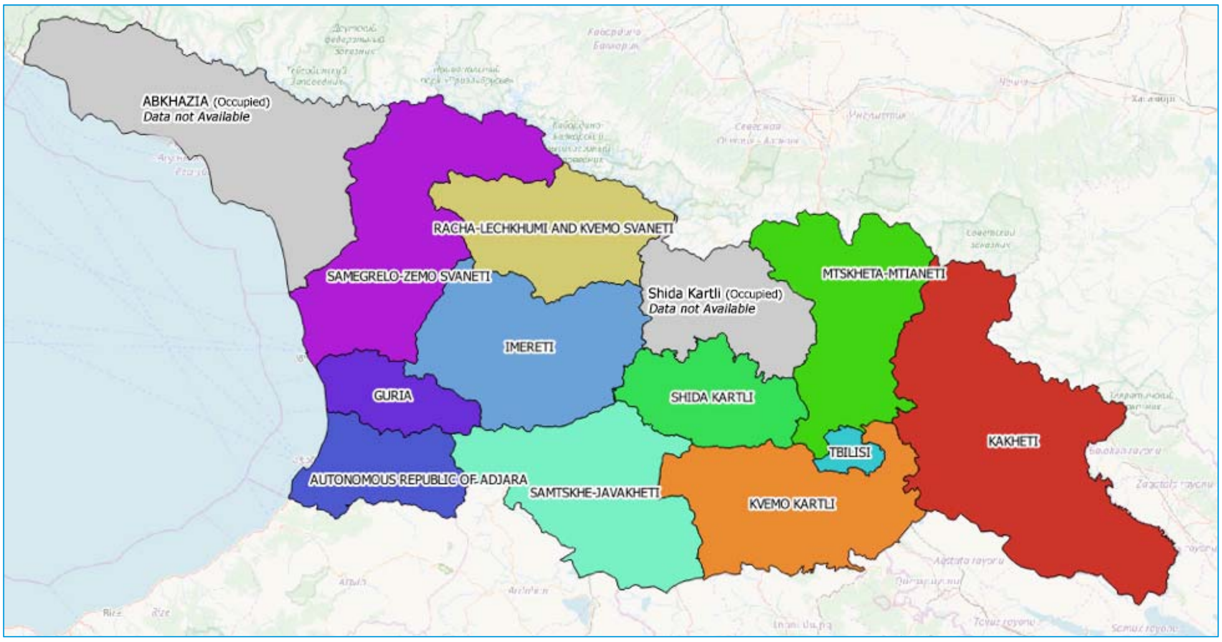
The computation of Tier II data has been performed entirely by GNCC. Tier I data from data providers have been deployed directly into GNCC servers, then processed from GNCC to get Tier II indicators. GNCC used an MSSQL database to store and process Tier I data. Access to Tier I data was not available to the data scientist.

5. RESULTS DERIVED FOR THE BIG DATA INDICATORS

According to the list of indicators (see Annex), 14 out of 15 indicators have been computed for Georgia. The check of all the indicators was revealed to be a difficult task. The main issue that led to such difficulties was related to Tier I data. GNCC's IT department found many inconsistencies and asked data providers to check and fix incorrect data; then, new data sets had to be ingested. This process took more than two weeks. The Tier II and Tier III computation process suffered from such a delay: not having access to Tier I data made it impossible for the data scientist to properly check the calculation process of each Tier II indicator. However, GNCC provided remarkable collaboration, working remotely with the data scientist to fix all the issues related to Tier II indicators. Such issues are reported in the corresponding indicator section.

Indicator BD10 has been dropped because of the lack of required Tier I data: the ISPs do not log traffic for fixed broadband contracts.

Figure 2: Regions, first-level local administrative areas (LAU1) of Georgia



5.1. BD01: PERCENTAGE OF THE TERRITORY AREA COVERED BY A MOBILE-CELLULAR NETWORK, BY TECHNOLOGY

This indicator is an extension of the official indicator with the aim of more in-depth geographical and technological breakdowns. The indicator should provide the understanding of access and spread of the mobile technology in different parts of the country. The resulting indicator should refer to the proportion of mobile-cellular coverage out of the territory in percent.

Figure 3: Percentage of the territory covered by a mobile-cellular network, by technology, for the whole country of Georgia

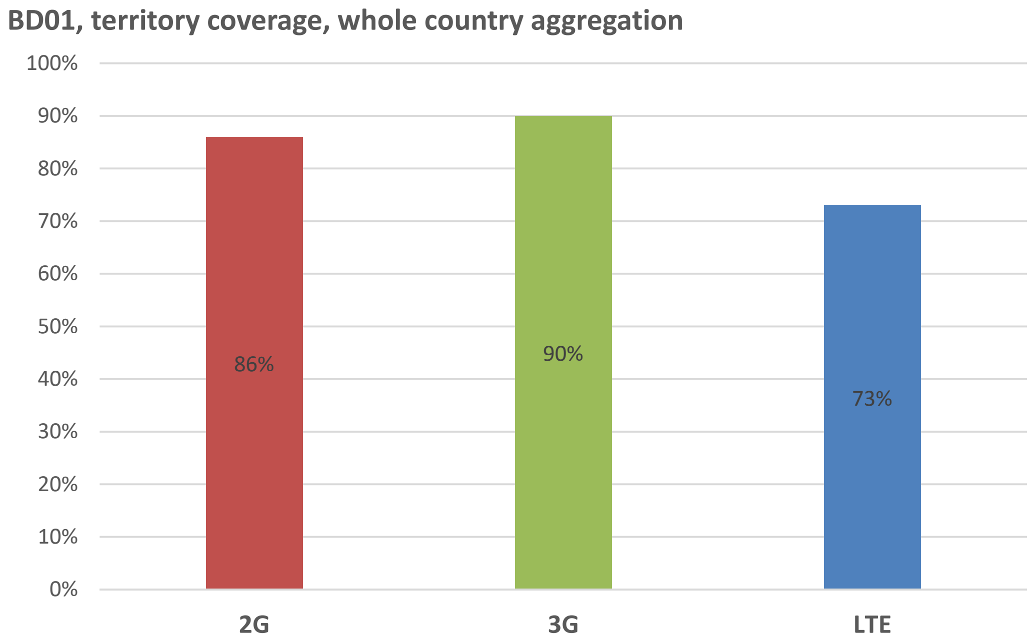


Figure 3 shows the percentage of the territory covered by a mobile-cellular network, for each type of technology: 2G and 3G networks cover approximately 90% territory of Georgia, while LTE covers a smaller percentage.

Georgia is a country covered by many mountains, rocks and a difficult terrains, therefore are non-coverage territories such as - The rocky part of Borjomi-Kharagauli national park and Egrisi range in western Georgia, The rocky part of Caucasus mountains in Samegrelo zemo Svaneti, Racha-lechkhumi, Mtskheta-Mtianeti and Kakheti, periglacial landscape in Samtskhe-Javakheti, mountainous and rocky territory between Kakheti and Mtsketa-Mtianeti (especially part of territory of Tusheti National park and Khevsureti) and other small areas with difficult terrain, the total area of which was ≈10%.

Details of the spatial coverage of 2G, 3G and LTE networks are provided respectively in figure 4, figure 5 and figure 6.

Figure 4: LAU2, Covered territory by a mobile-cellular 2G network

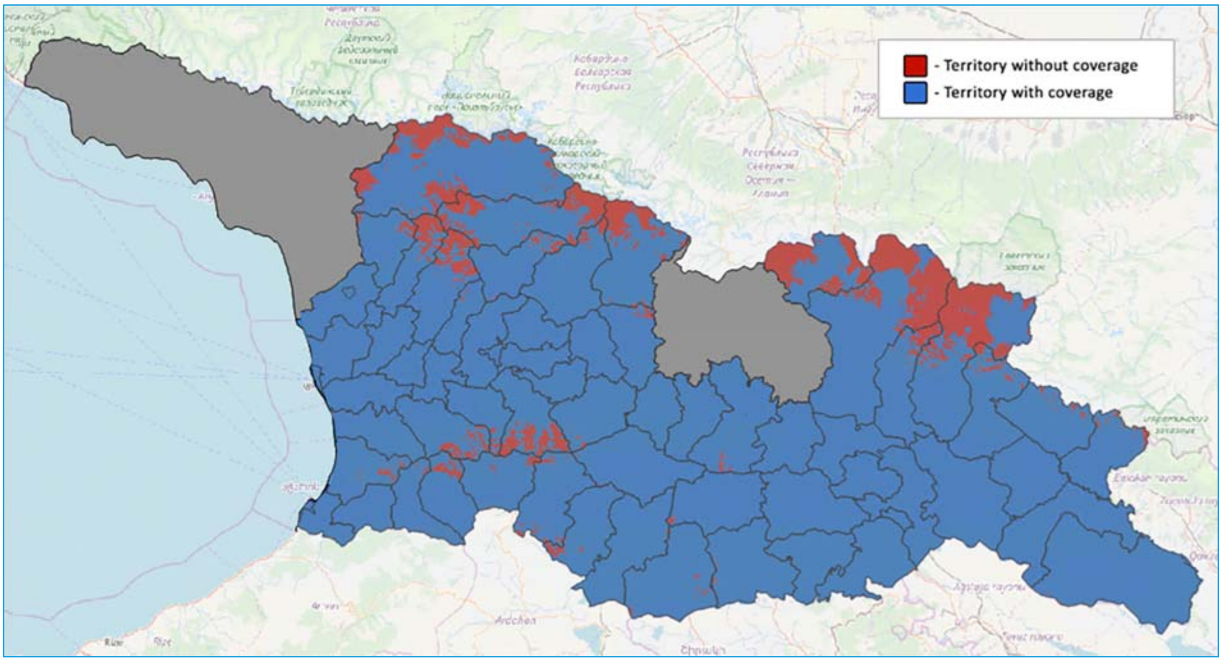


Figure 5: LAU2, Covered territory by a mobile-cellular 3G network

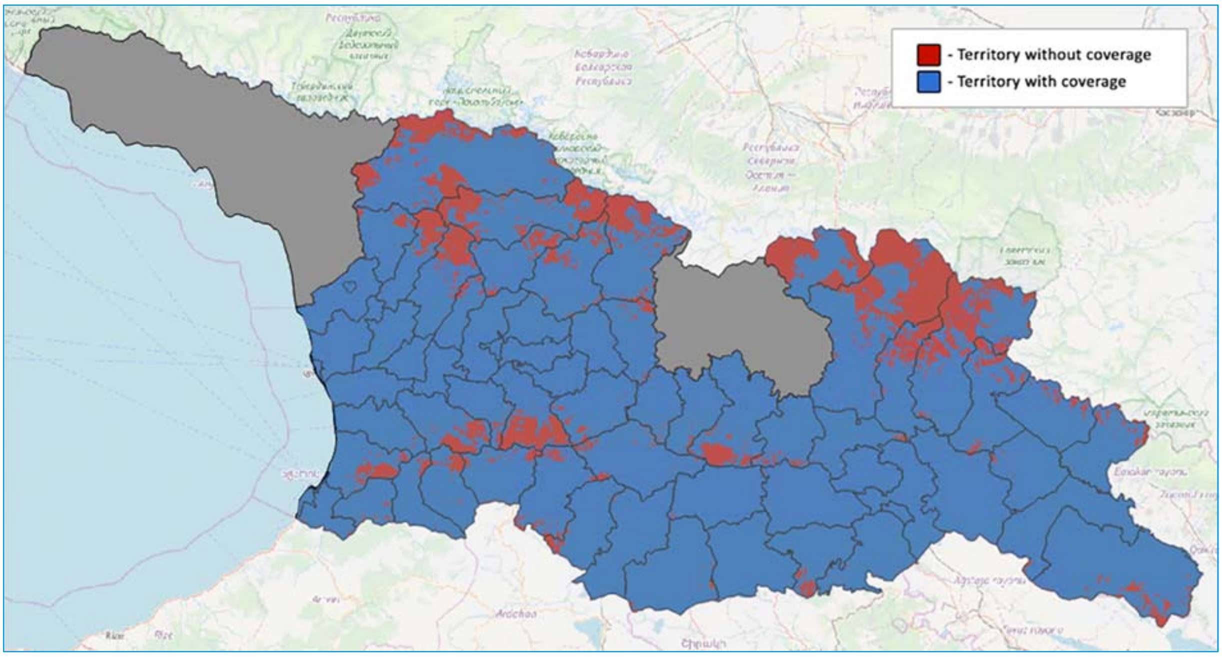
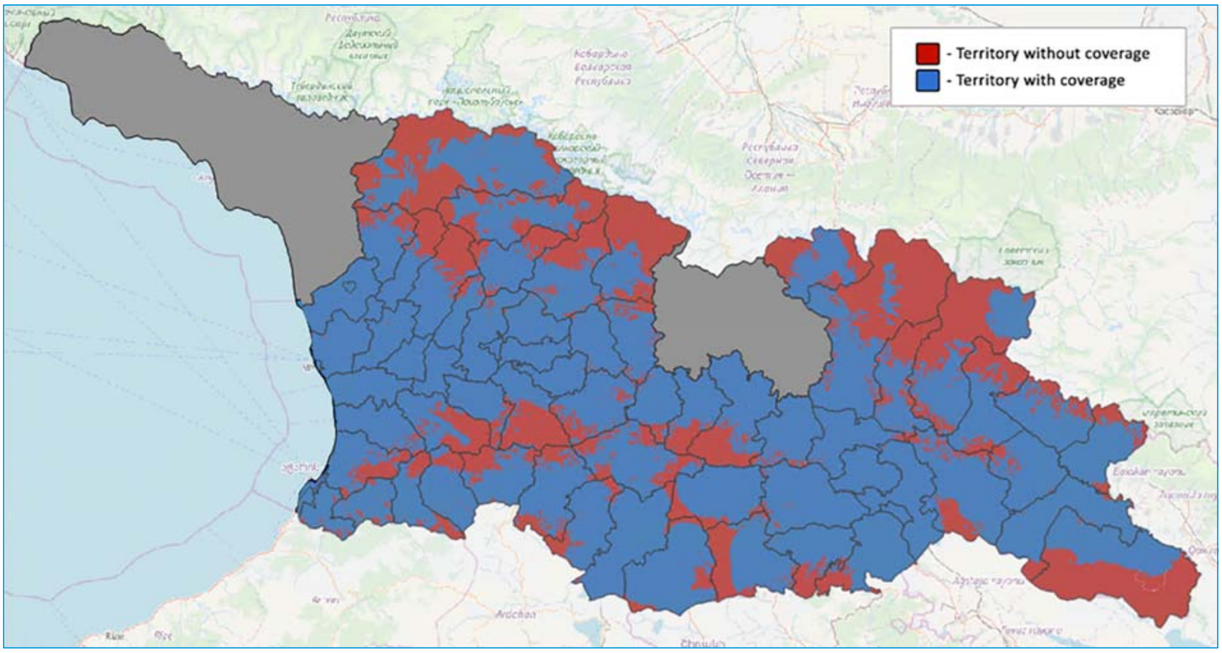


Figure 6: LAU2, Covered territory by a mobile-cellular LTE network



5.2. BD02: PERCENTAGE OF THE POPULATION COVERED BY A MOBILE-CELLULAR NETWORK, BY TECHNOLOGY

This indicator is an extension of the official indicator with the aim of more in-depth geographical and technological breakdowns. The indicator should provide the understanding of access and spread of the mobile technology in different parts of the country. The resulting indicator should refer to the proportion of inhabitants within range of a mobile-cellular signal out of the total number of inhabitants in the specific area in per cent. This is an extension of previous indicator (BD01) on land area coverage and is proposed to calculate at the same time with, simply by including the population calculation to the algorithm. The source data and the methodology for calculation is basically the same.

Figure 7: Percentage of the population covered by a mobile-cellular network, by technology, for the whole country of Georgia

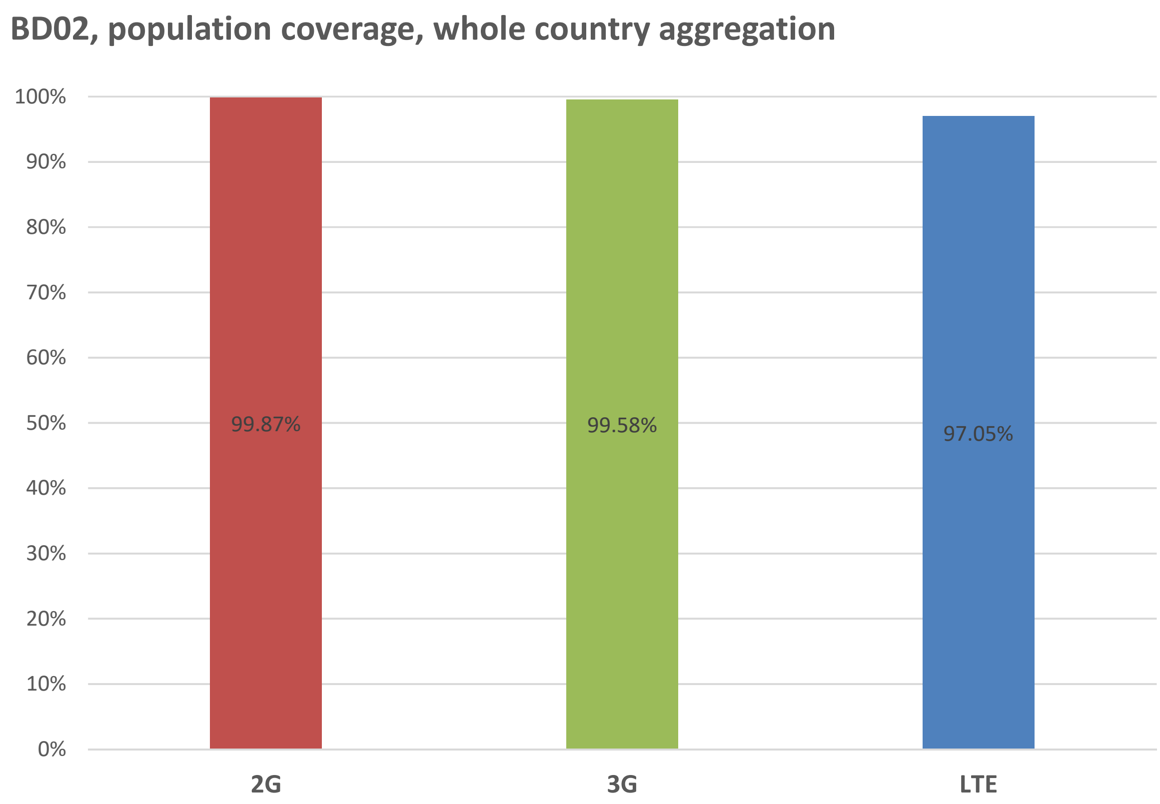


Figure 7 shows the percentage of the population covered by a mobile-cellular network, for each type of technology: 2G and 3G networks cover approximately 99% of Georgia's population, while LTE covers a slightly smaller percentage. Approximately 30% of the Georgian population lives in the Tbilisi region, which is completely covered by 2G, 3G and LTE service. Other regions are less covered: however, in most regions, all three technologies have a similar coverage percentage (see figure 8).

Details of the spatial coverage of 2G, 3G and LTE networks are provided respectively in figure 9, figure 10 and figure 11. Note that color of each 1 km^2 area is based on whether or not this area is covered.

Figure 8: Percentage of the population covered by a mobile-cellular network, by technology, for each administrative LAU1 region of Georgia

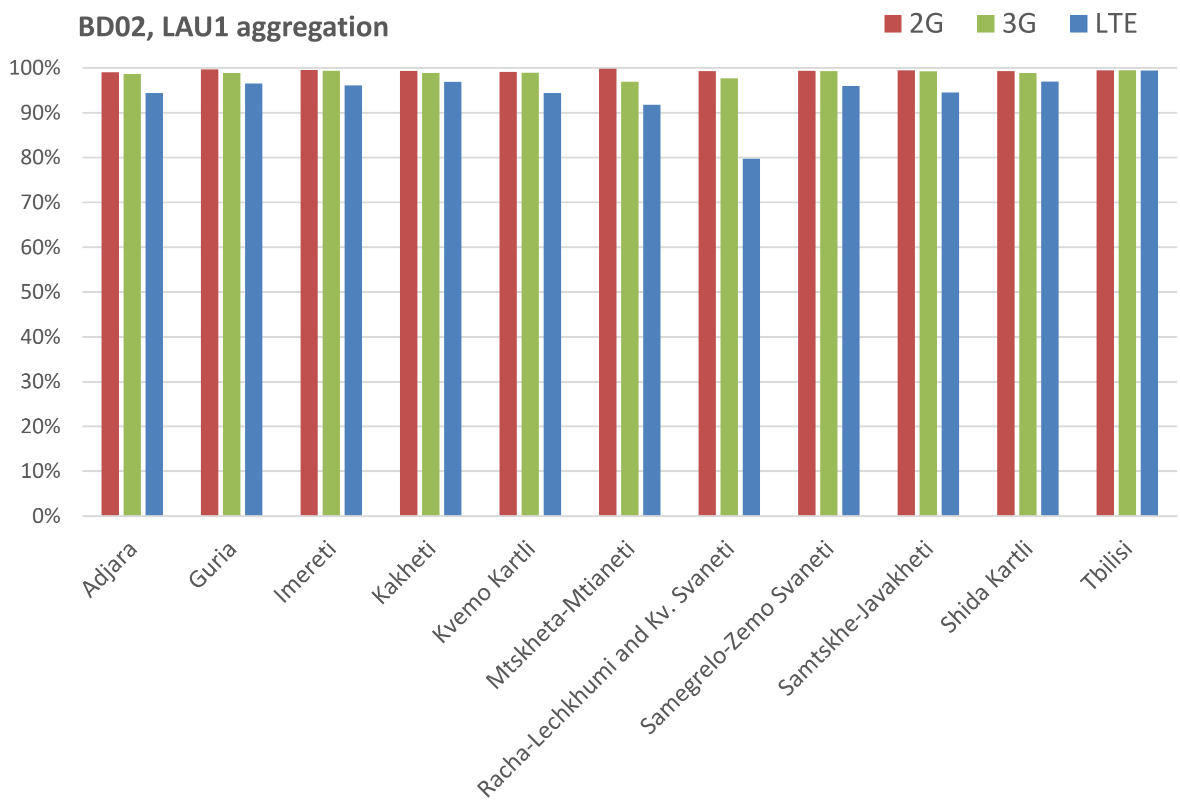


Figure 9: Covered population by a mobile-cellular 2G network in every 1 km^2

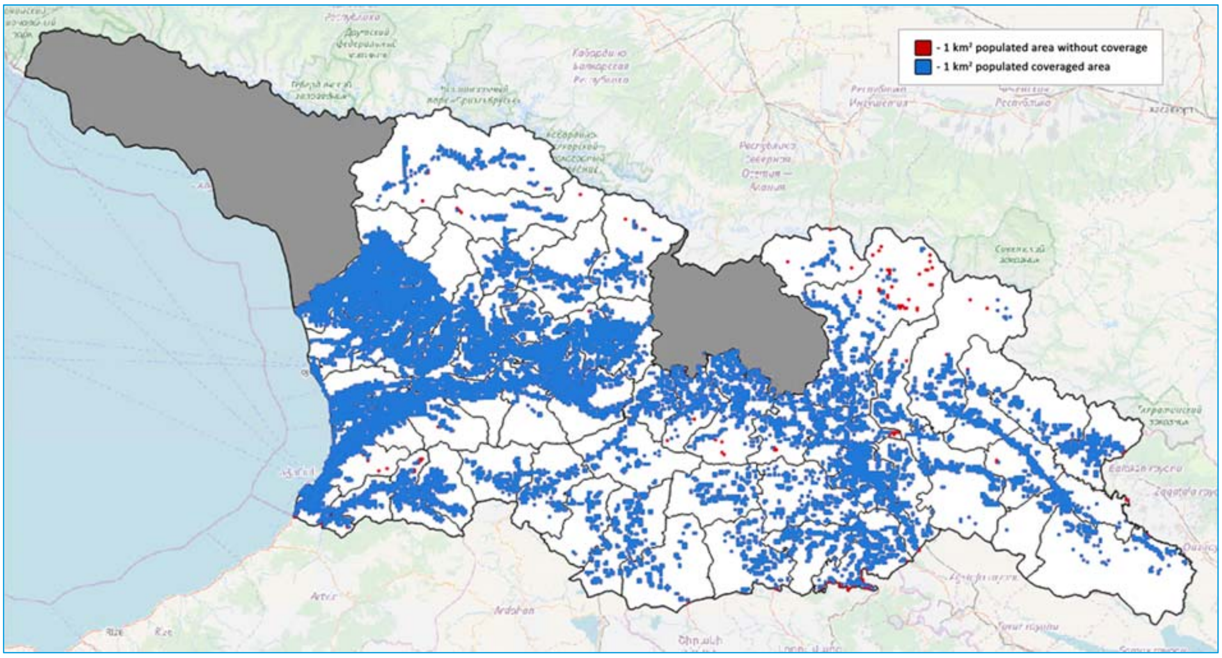


Figure 10: Covered population by a mobile-cellular 3G network in every 1 km^2

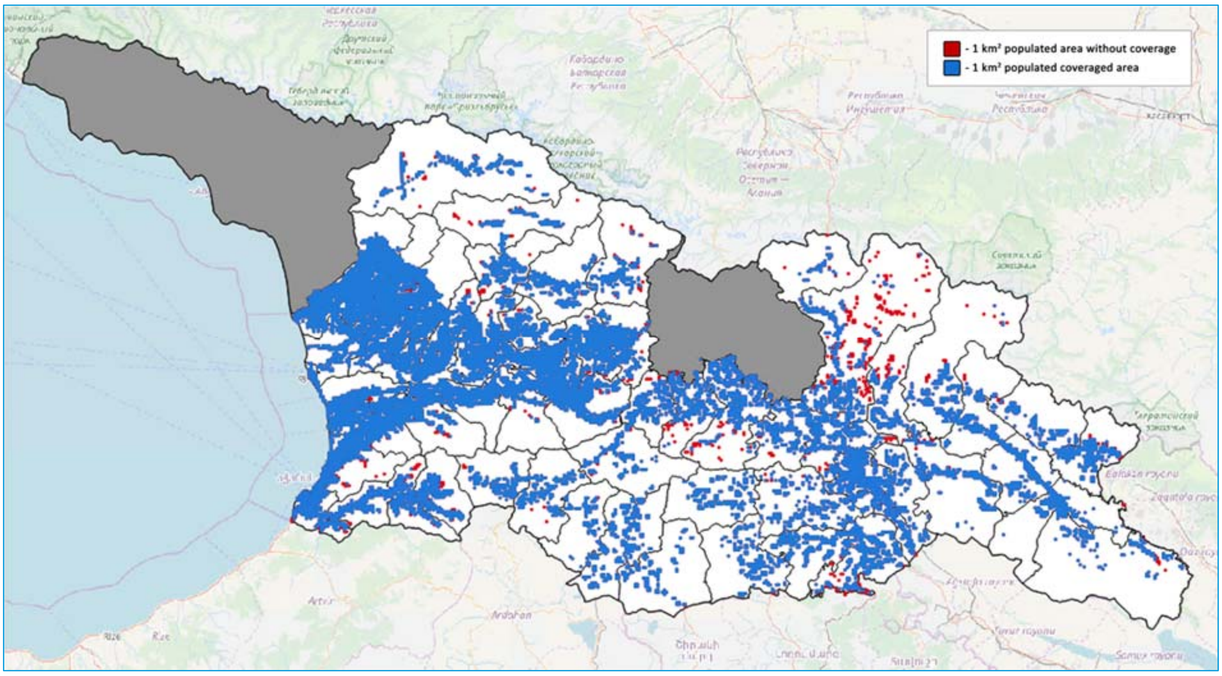


Figure 11: Covered population by a mobile-cellular LTE network in 1 km²

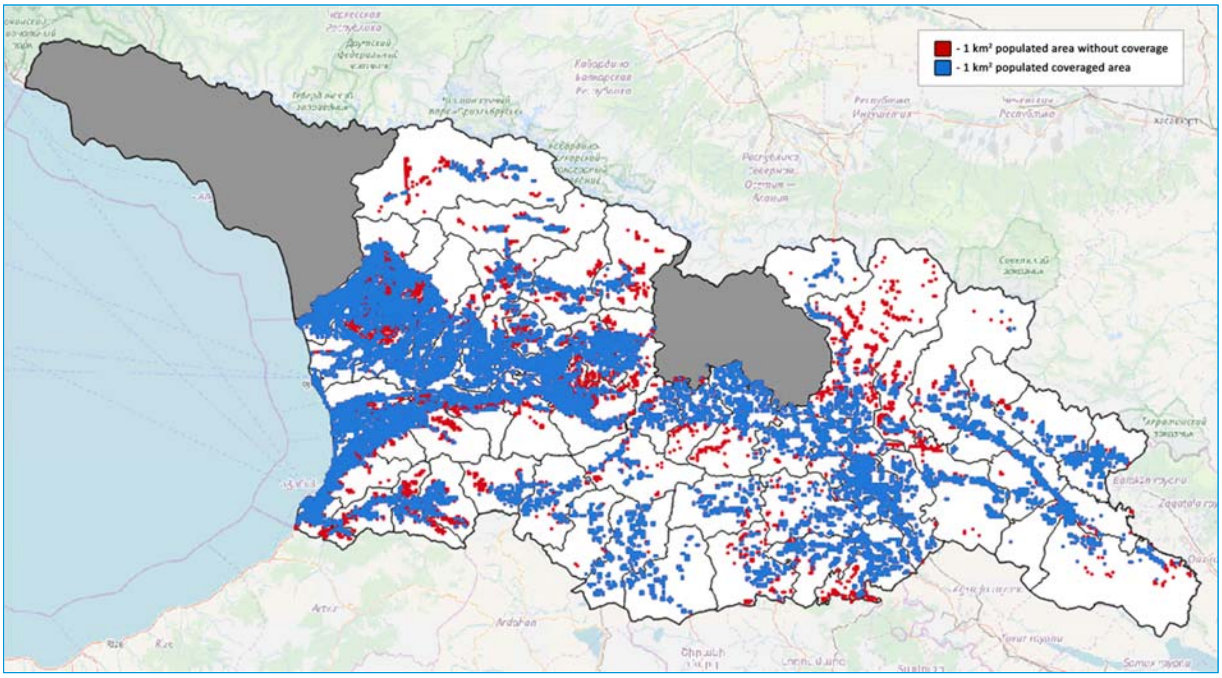


Figure 12: Covered population by a mobile-cellular 2G network, LAU2

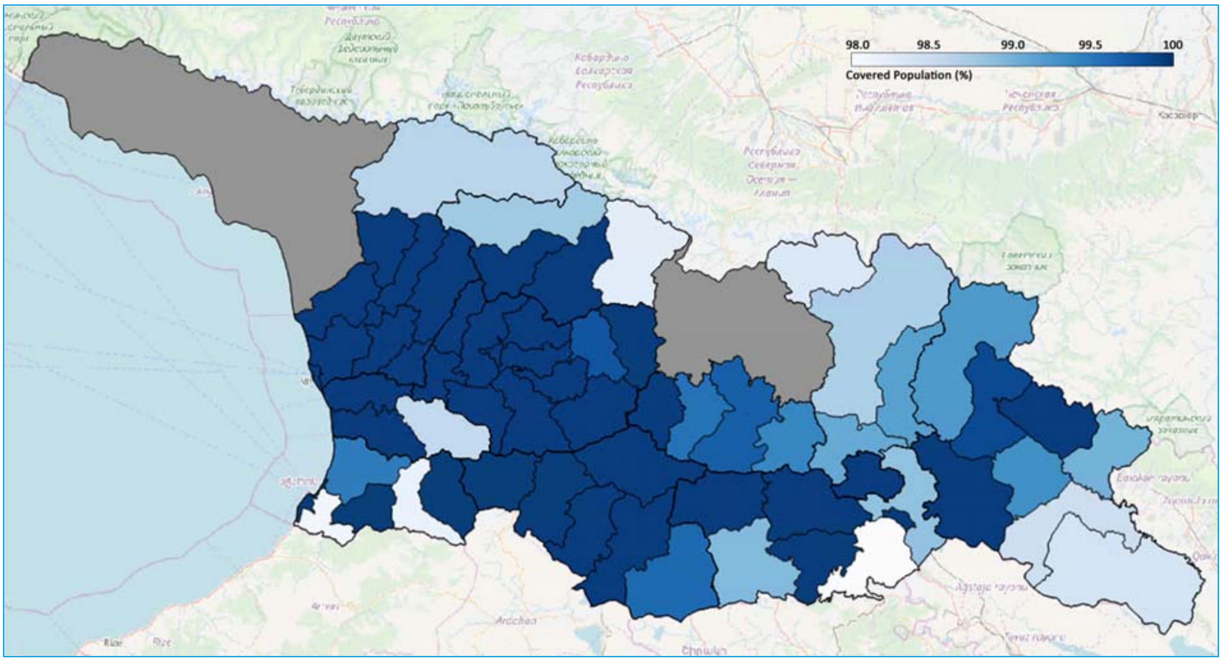


Figure 13: Covered population by a mobile-cellular 3G network, LAU2

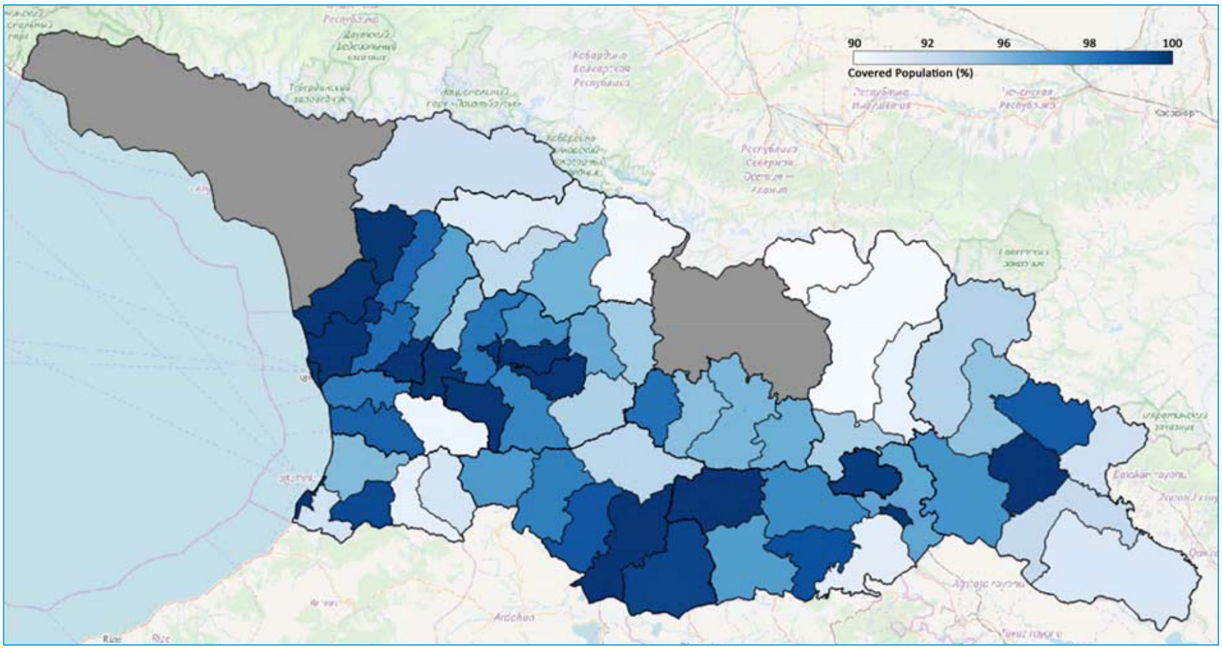
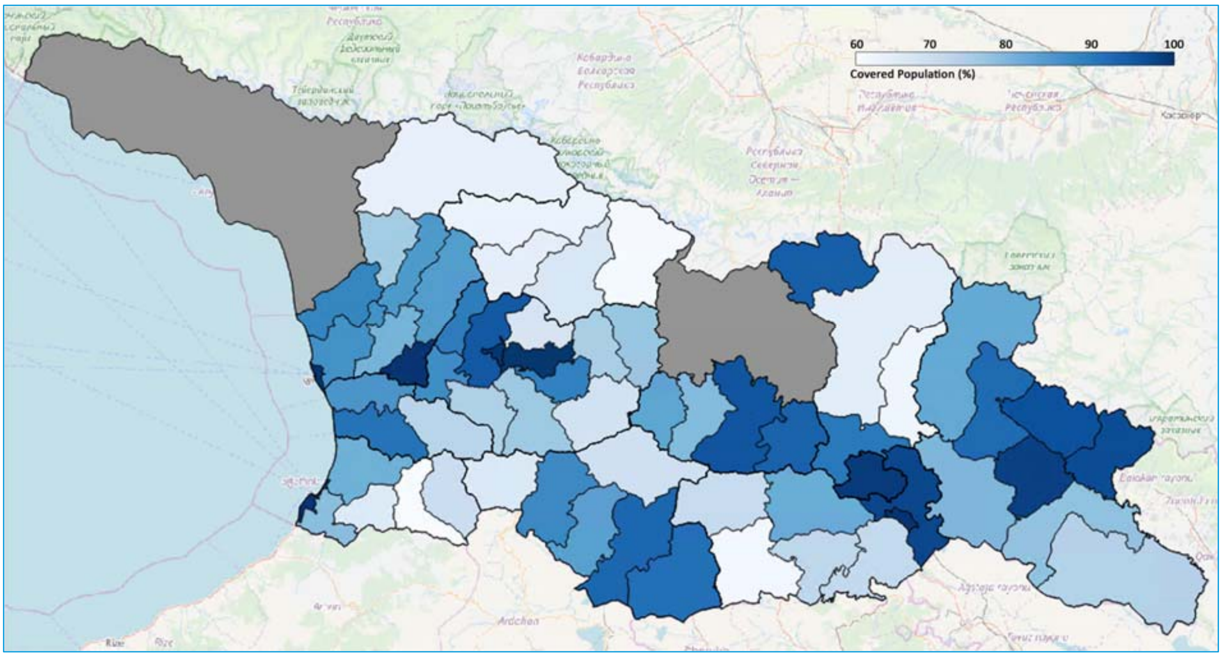


Figure 14: Covered population by a mobile-cellular LTE network, LAU2



5.3. 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 mobile voice calling 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 counted as a fact of the usage of the 2G network, and one incoming SMS via 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 with 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).

Observing the distribution for the whole country (figure 15), it is clear how the most used networks for non-IP-related activities are 2G, followed by 3G. Such a distribution is stable even when aggregating according to LAU1 regions (figure 16), with a slightly higher usage of LTE for the Tbilisi region. In figure 17, figure 18 and figure 19, the percentage of usage of, respectively, the 2G, 3G and LTE networks are aggregated on a LAU2 basis. As expected after reading results with LAU1 aggregation, network usage is quite uniform across the whole country, with some outlier for the LTE network.

Figure 15: Usage of mobile-cellular networks for non-IP-related activities, by technology, for the whole country of Georgia

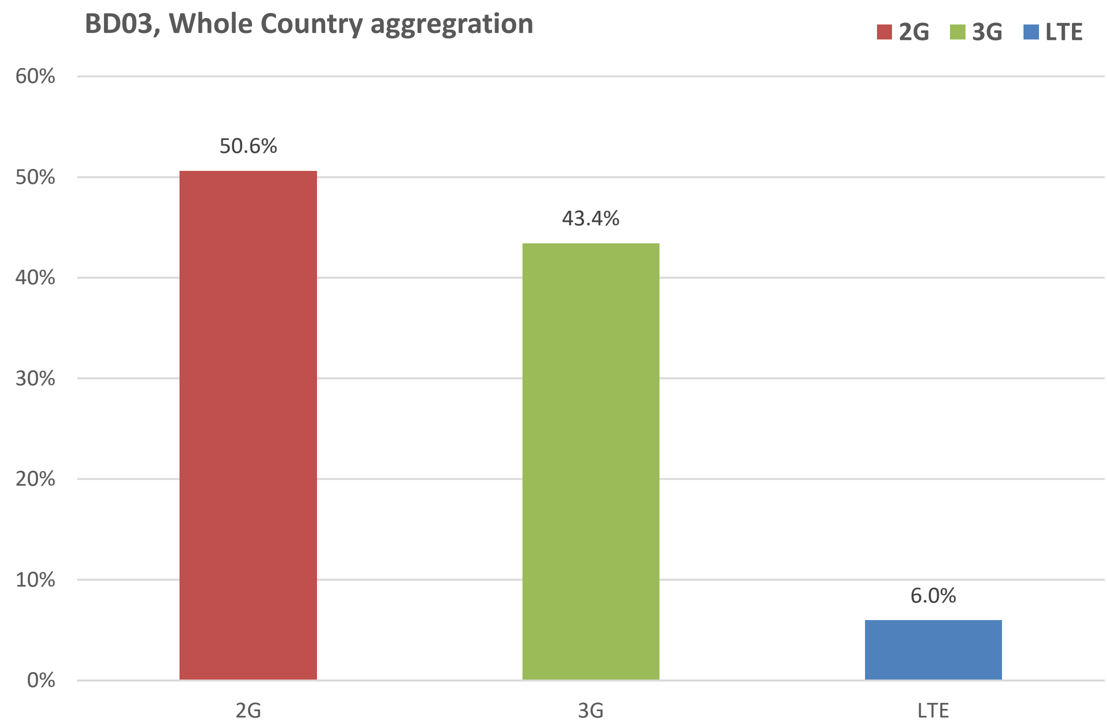


Figure 16: Usage of mobile-cellular networks for non-IP-related activities, by technology (LAU1 regions)

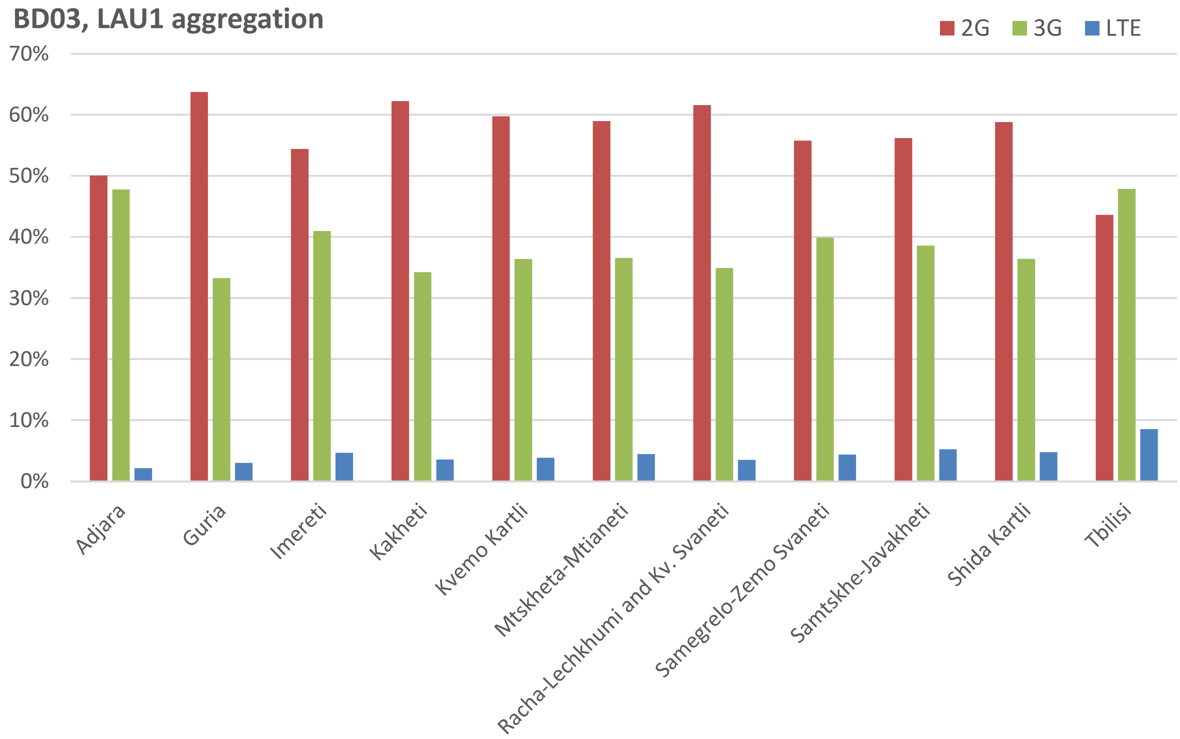


Figure 17: Usage of mobile-cellular 2G networks for non-IP-related activities in LAU2 districts

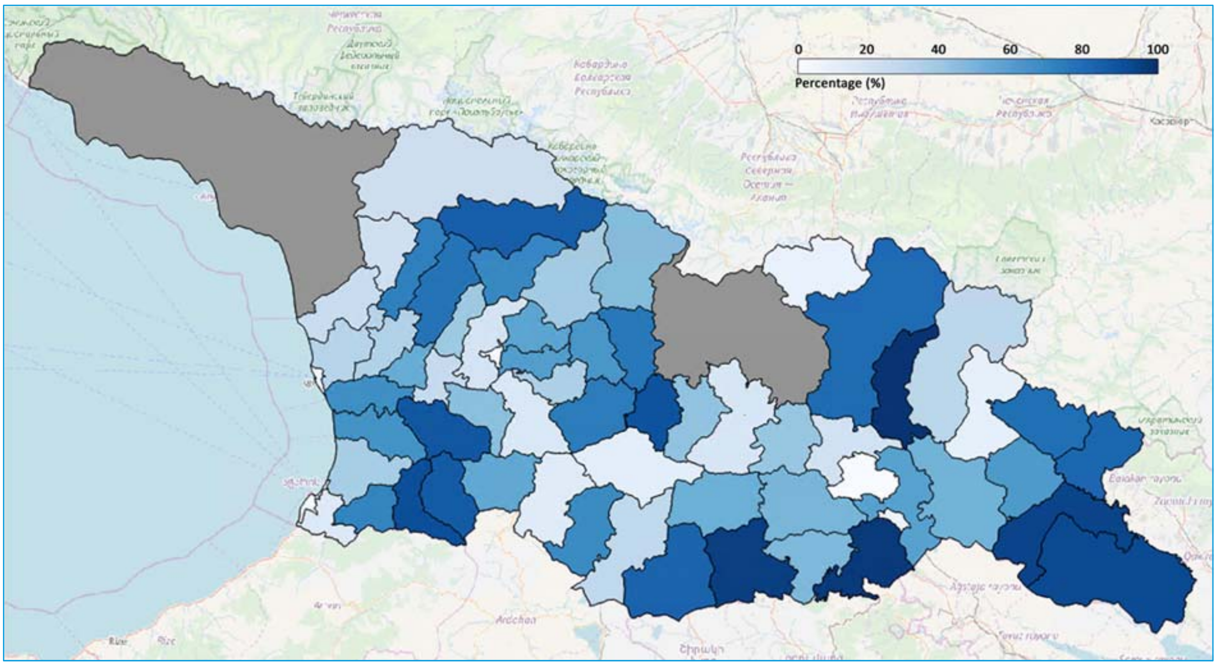


Figure 18: Usage of mobile-cellular 3G networks for non-IP-related activities in LAU2 districts

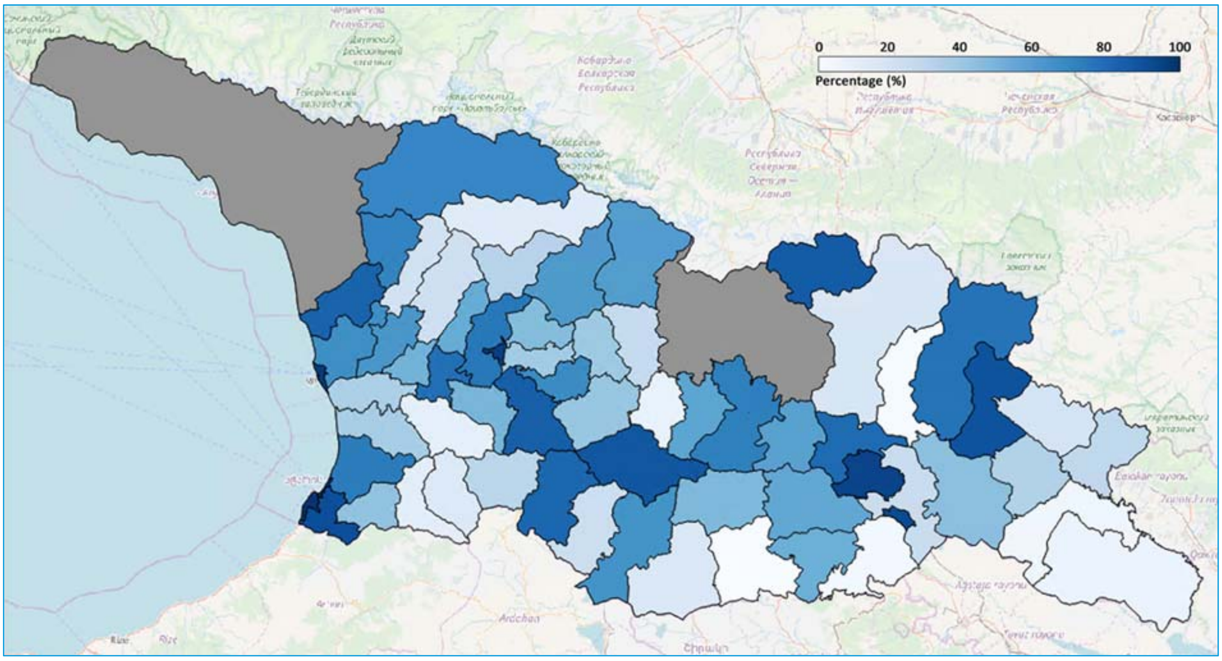
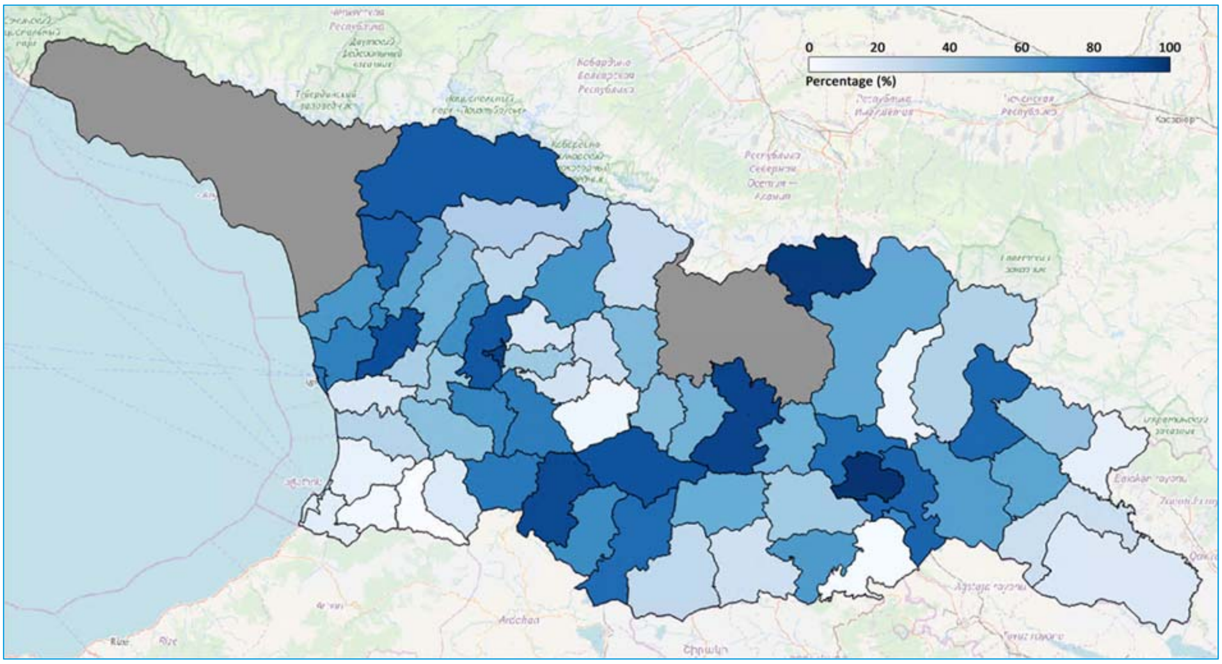


Figure 19: Usage of mobile-cellular LTE networks for non-IP-related activities in LAU2 districts



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

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

Conversely, to non-IP-related activities, mobile Internet access is mainly exploiting the LTE network. In Figure 20, it is clear how the LTE network represented the most used network for mobile internet access, for the whole country of Georgia. It is similar when the indicator is broken down by LAU1 regions: apart from the region of Kakheti, Racha-Lechkhumi, Kv. Svaneti and Samtskhe-Javakheti, LTE is the most-used network for Internet access (figure 21). In figure 22, figure 23 and figure 24, the usage of, respectively, the 2G, 3G and LTE networks is provided for each LAU2 districts.

Figure 20: Distribution of network usage for Internet access, by technology, for the whole country of Georgia

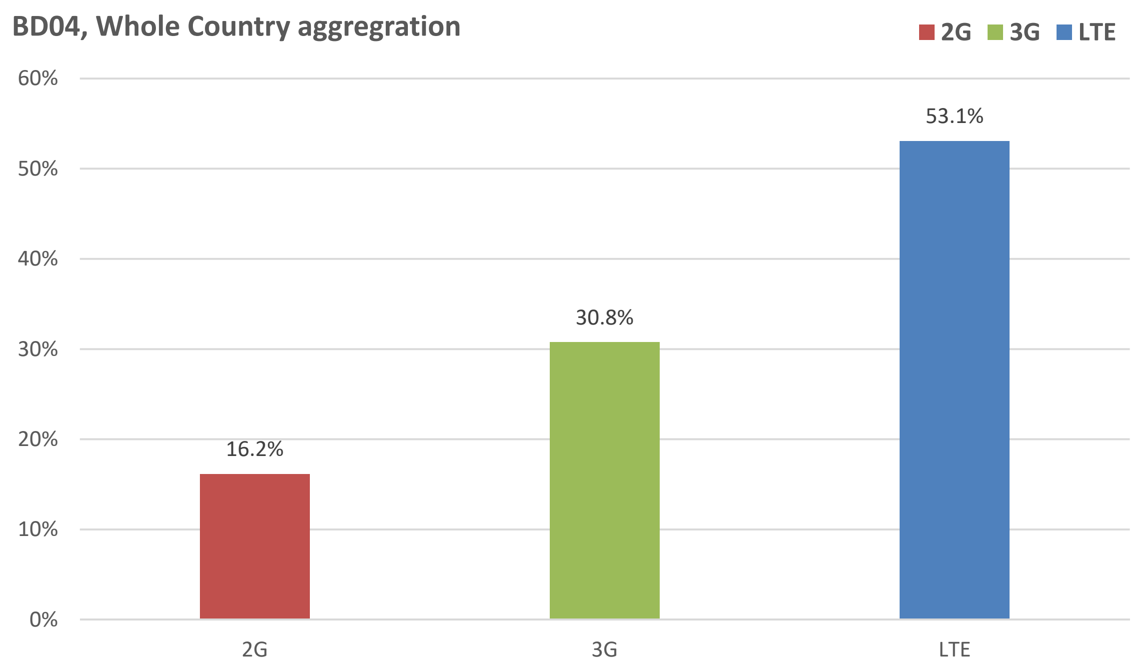


Figure 21: Distribution of network usage for Internet access, by technology, in LAU1 regions

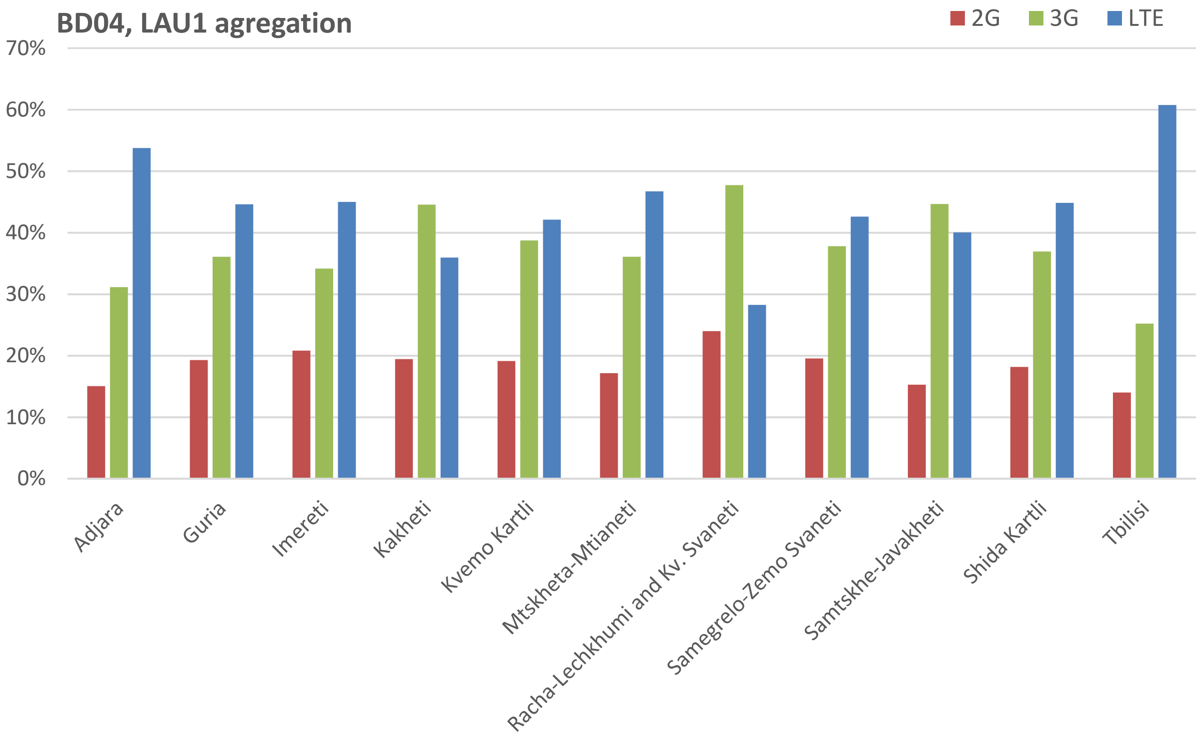


Figure 22: Usage of mobile-cellular 2G networks for IP-related activities in LAU2 districts

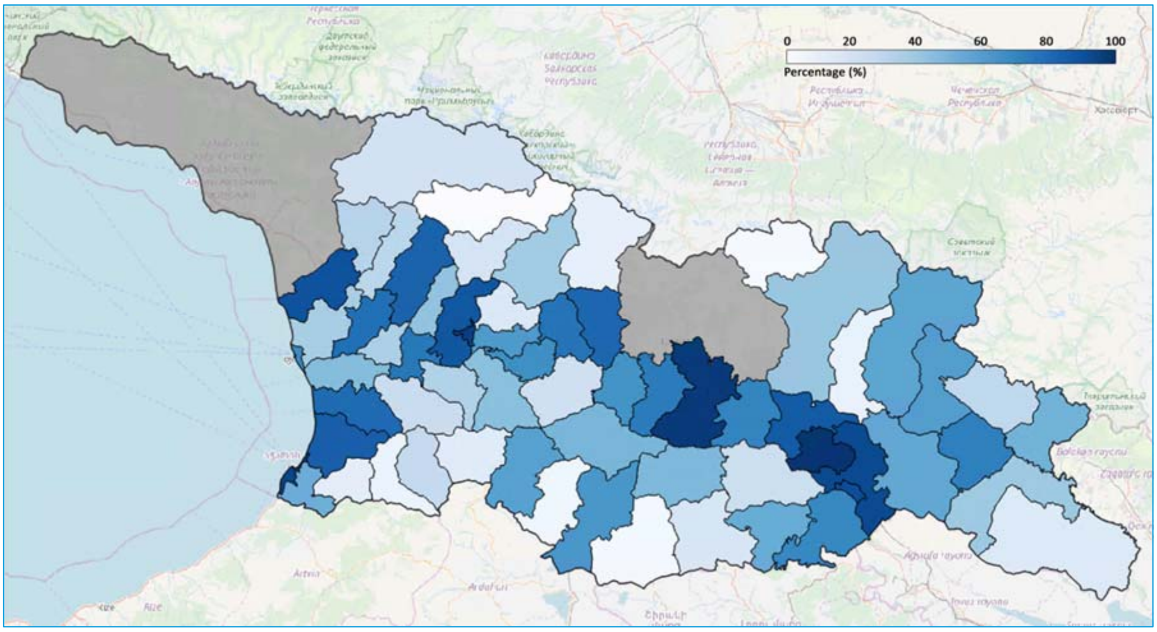


Figure 23: Usage of mobile-cellular 3G networks for IP-related activities in LAU2 districts

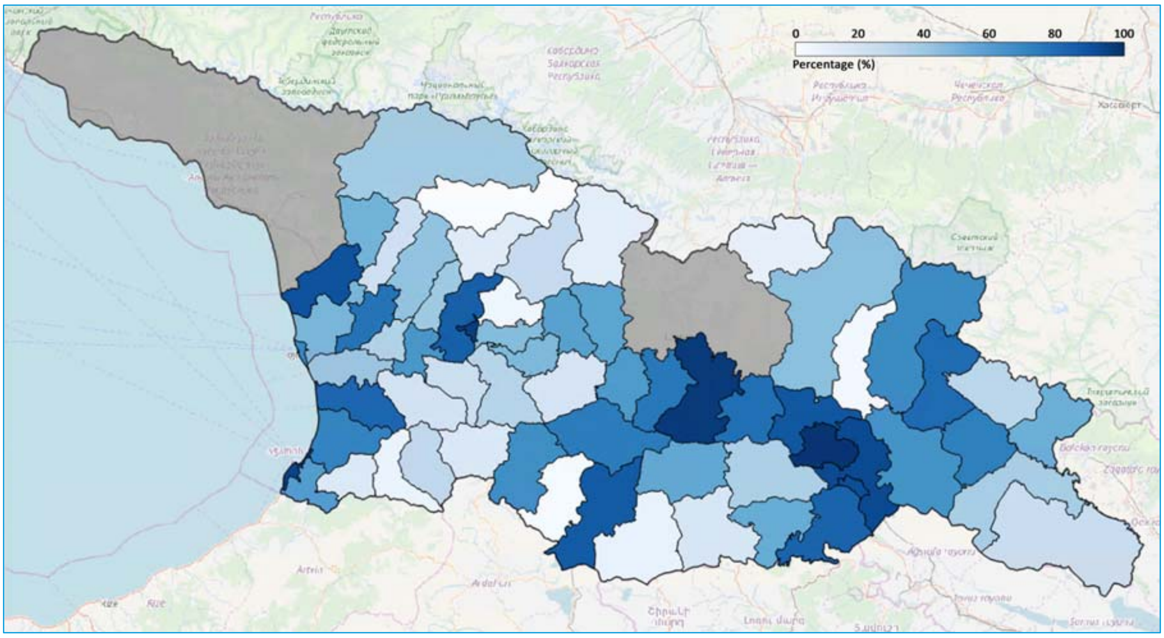
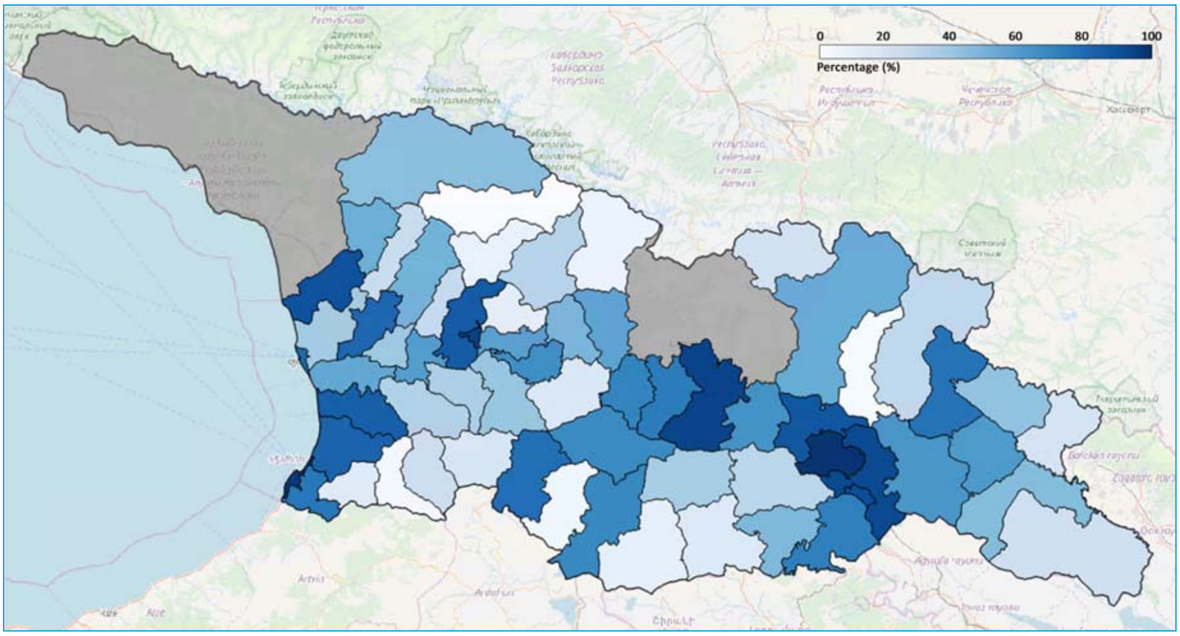


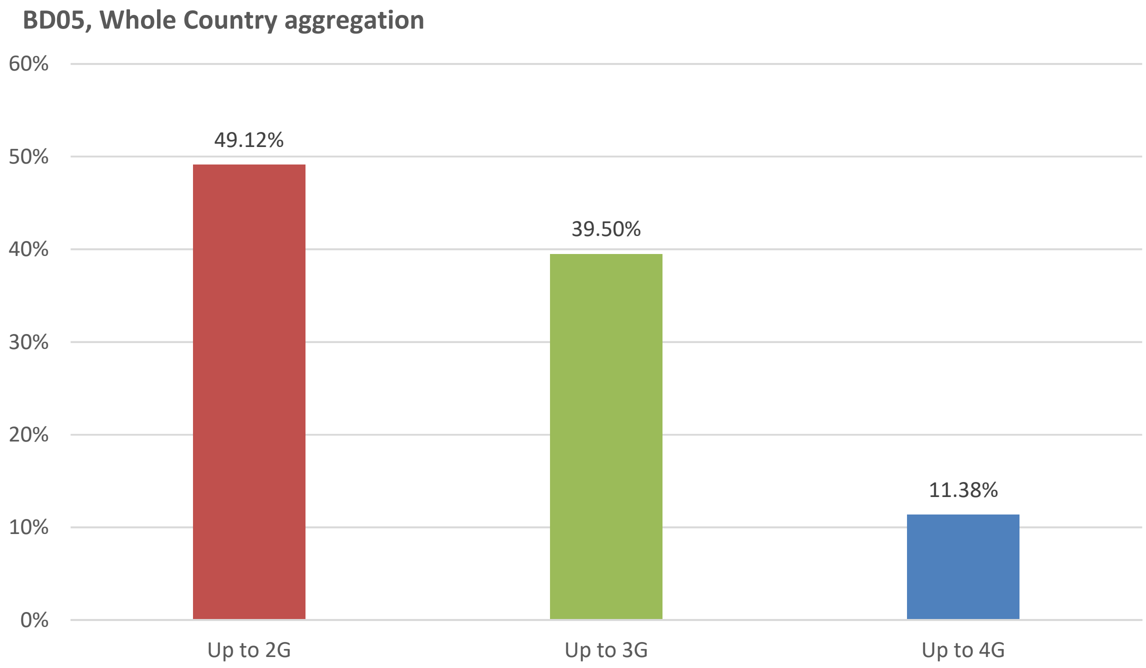
Figure 24: Usage of mobile-cellular LTE networks for IP-related activities in LAU2 districts



5.5. BD05: NUMBER OF SUBSCRIPTIONS WITH ACCESS TO TECHNOLOGY

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

Figure 25: Proportion of access to 2G, 3G and LTE technologies, by subscriptions, in the whole of Georgia



In figure 25, the distribution of subscriptions for each technology is provided, aggregated for the whole country of Georgia. Most of the population of Georgia has only accessed 2G networks, while just a small percentage (< 12%) has accessed the LTE network. Given this, and recalling results from the BD04 indicator, it suggests that Internet access is mostly exploited with the LTE network, but the majority of the subscribers either do not use the Internet over mobile devices. The existence of subscribers with unknown access could be explained by the different time periods of the two data sets used to calculate geographic coverage. The CDR and IPDR represented the last quarter of 2016, while the geographic information system coverage data was from March 2017, when GNCC was made aware of the removal of some antennas, although it was not specified which and where. In figure 26, the distribution of subscriptions for each LAU1 region shows how 2G is the only accessed network for most of population across all regions, Tbilisi included. Again, this may be caused by subscribers not using high-speed Internet with their devices. Spatial distributions according to LAU3 communities/towns are provided in figure 27, figure 28 and figure 29.

Figure 26: Proportion of access to 2G, 3G and LTE technologies, by subscriptions, in LAU1 regions

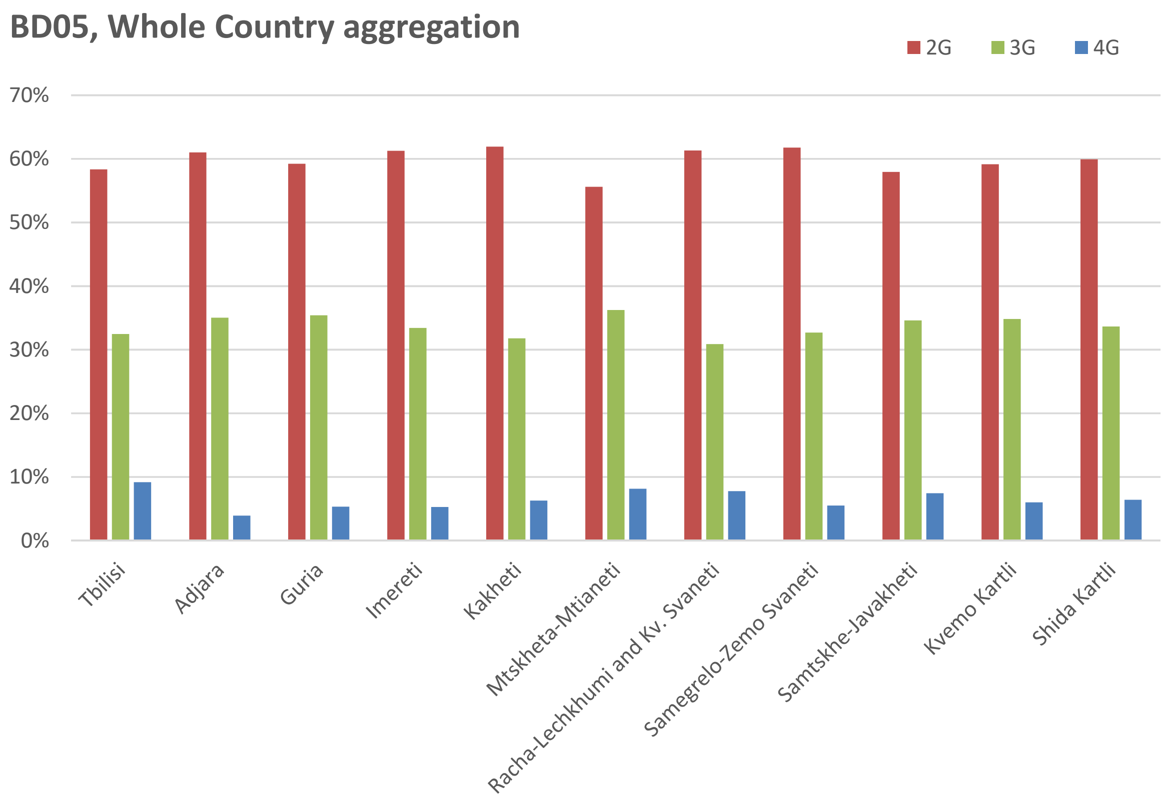


Figure 27: Proportion of subscribers that have access only up to 2G networks in LAU2 districts

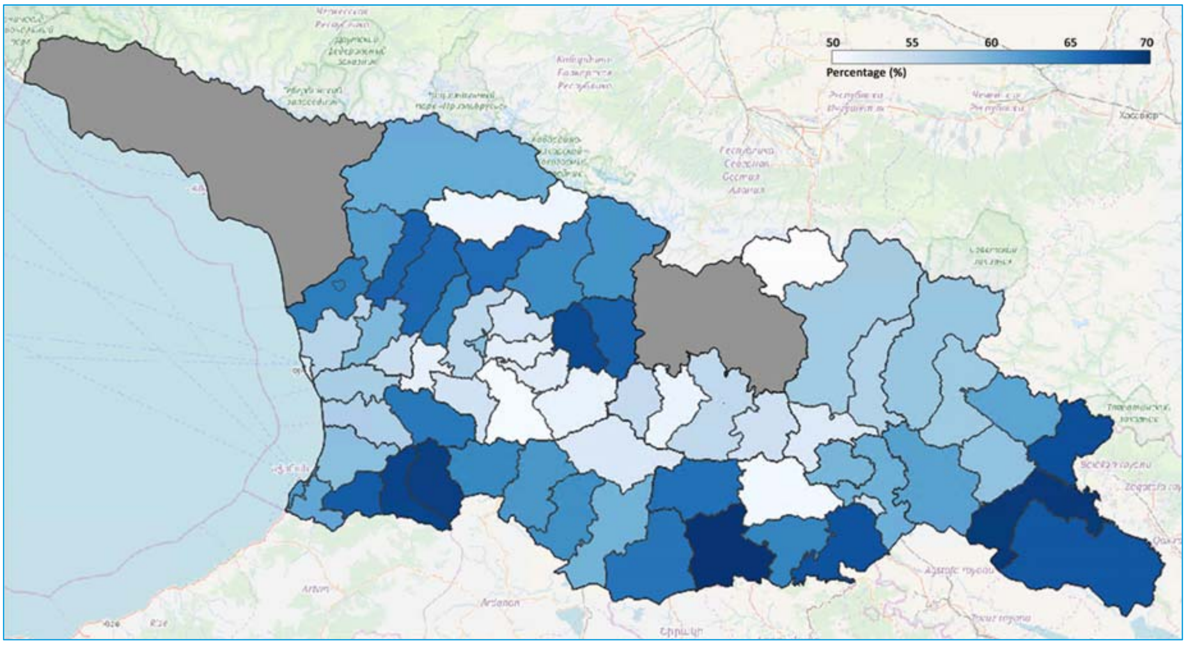


Figure 28: Proportion of subscribers that have access only up to 3G networks in LAU2 districts

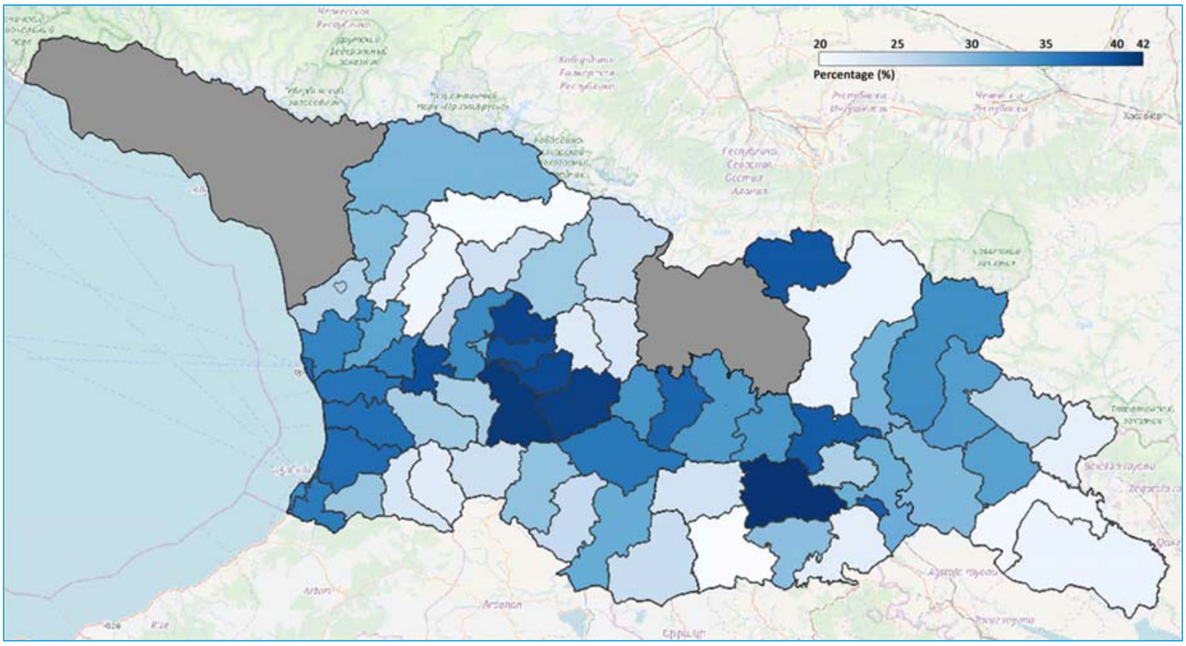
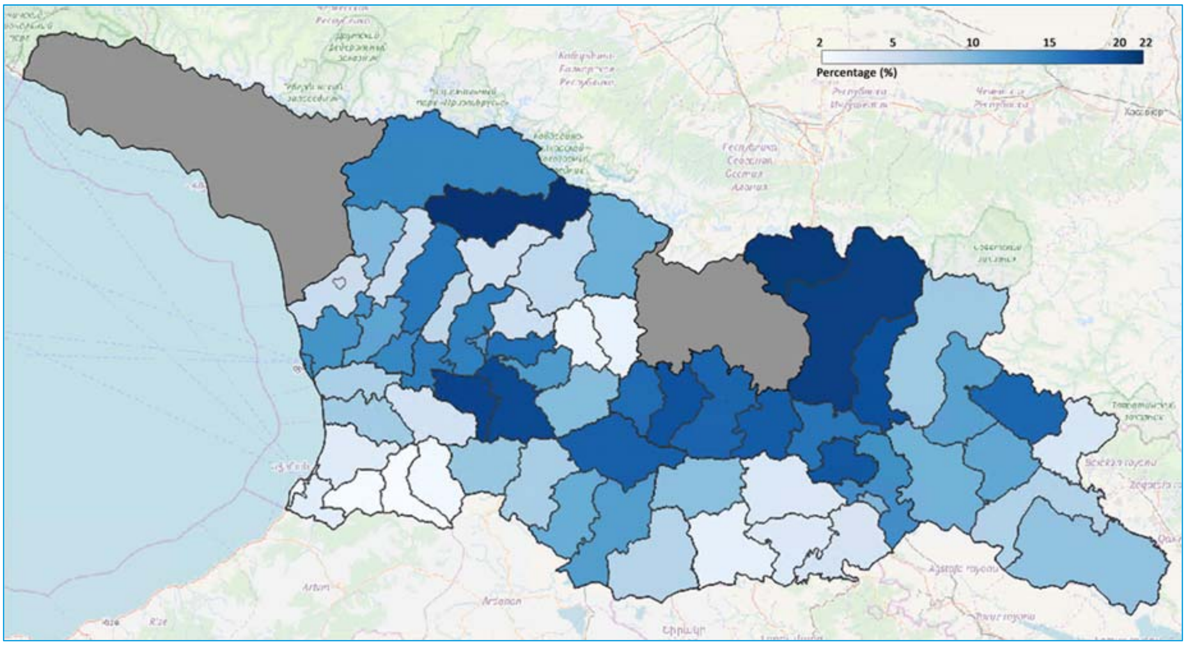


Figure 29: Proportion of subscribers that have access only up to LTE networks in LAU2 districts



5.6. BD06: ACTIVE MOBILE VOICE AND BROADBAND SUBSCRIPTIONS, BY EVENT TYPE

This indicator refers to the number of active mobile subscriptions disaggregated by type of contract (commercial, private; prepaid, post-paid; no-data, voice and data, data only). The total number of mobile voice and broadband subscriptions in Georgia was 4 980 200. In source data was not possible identify contract types and therefore there was used event type. Accordingly this, the distribution of subscriptions for both event types, aggregated for the whole country, is provided in figure 30. LAU1 aggregation (figure 31) highlights the same trend observed with whole country aggregation, in terms of voice/data relative number of subscriptions. In figure 32 and figure 33 show the spatial distribution of subscriptions for each type of event. Note that the same scale is used for both maps, in order to highlight the absolute difference between the two event types.

Figure 30: Number of active subscriptions by event type for the whole country of Georgia

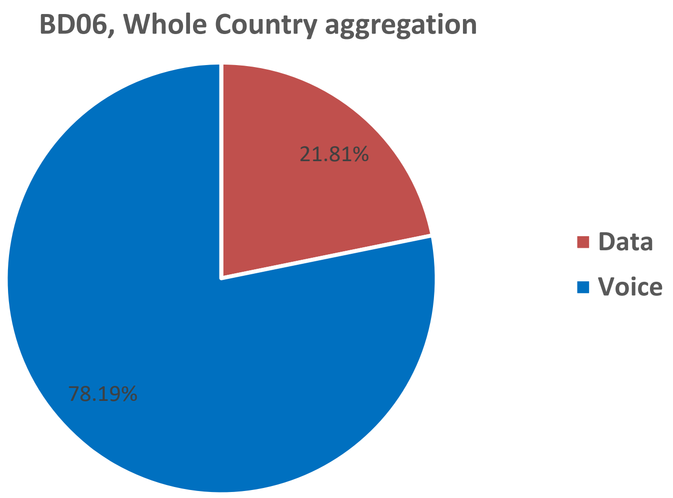


Figure 31: Number of subscriptions for each administrative region by type of event (voice and data)

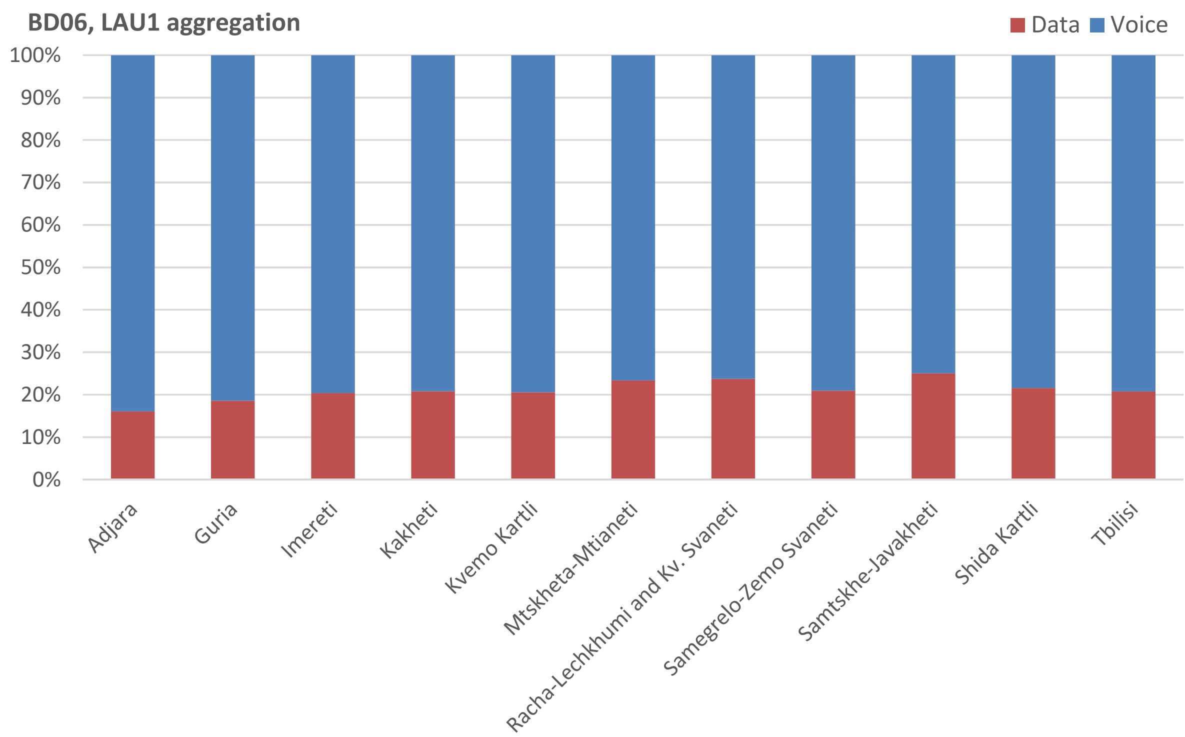


Figure 32: Active subscriptions for voice events, LAU2 aggregation

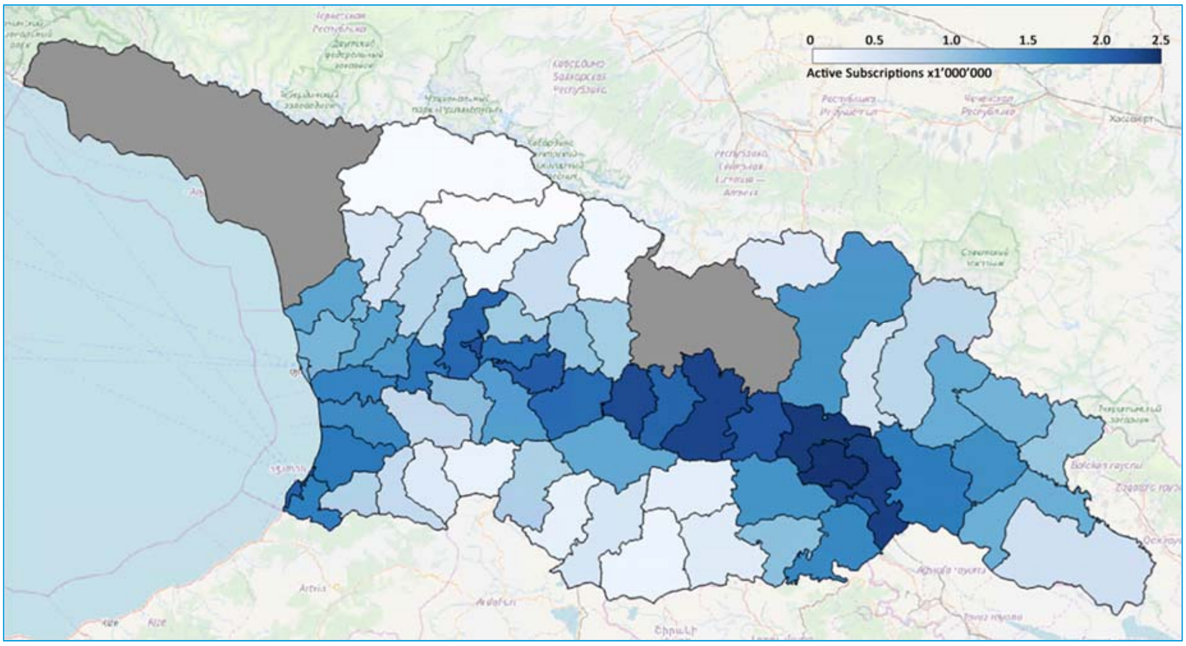
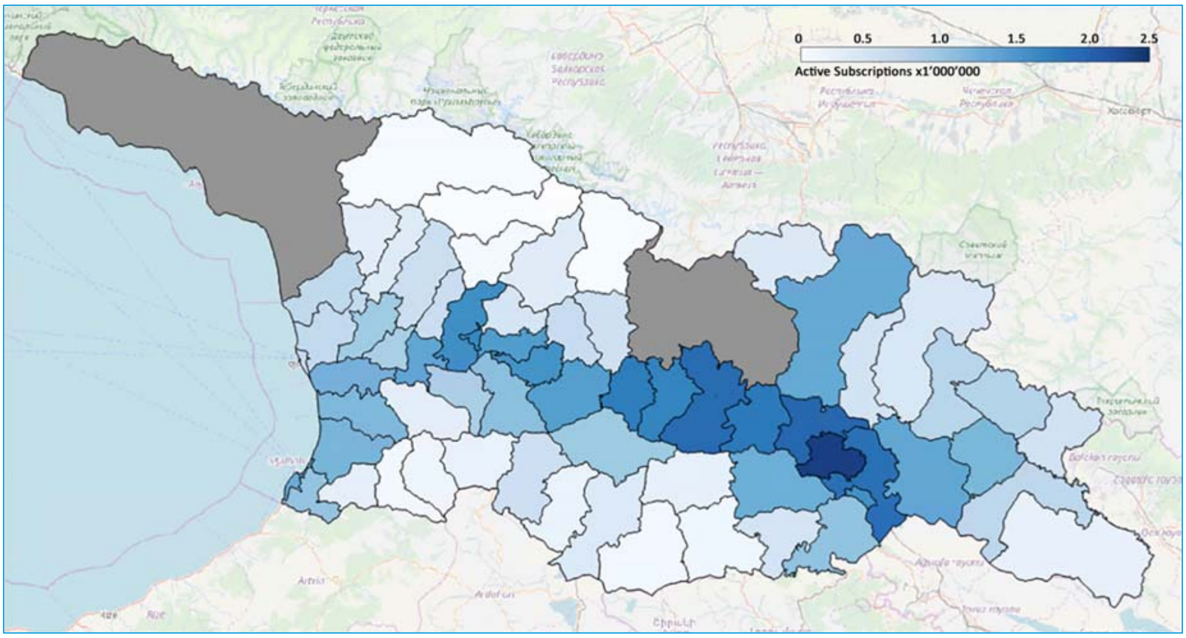


Figure 33: Active subscriptions for data events, LAU2 aggregation



5.7. BD07: AVERAGE NUMBER OF ACTIVE MOBILE SUBSCRIPTIONS PER DAY, BY EVENT TYPE

This indicator refers to the average number of active mobile subscriptions on a daily basis disaggregated by the type of the event. Compared to indicator BD06, this indicator refers to the daily activity of the subscribers and indicates the proportion of active use of the mobile voice and messaging service, and Internet access via mobile networks. The indicator shows how many subscriptions are active on a daily basis during the observed period. The higher this number is, the more subscriptions are using and accessing the mobile network on a daily basis. The ideal number should be equal to the indicator from BD06, meaning that all subscriptions are active and using the mobile network on all the days of the observed period.

From BD06, it can be seen that there are 4 980 200 total subscribers based on the data received from the data providers. The average number of active subscribers per day represents 58.27% of the total, meaning that more than one-third of the subscribers do not actually use a mobile device on a daily basis. One of the reasons why this number is not 100% is that penetration is 148% which means that some subscribers have more than one sim card, and they do not use the extra sim card on daily basis. However, there are always some subscribers who do not use the phone every day, and this indicator shows the proportion of such subscribers.

Figure 34: Total number of subscribers and average active subscribers per day in Georgia

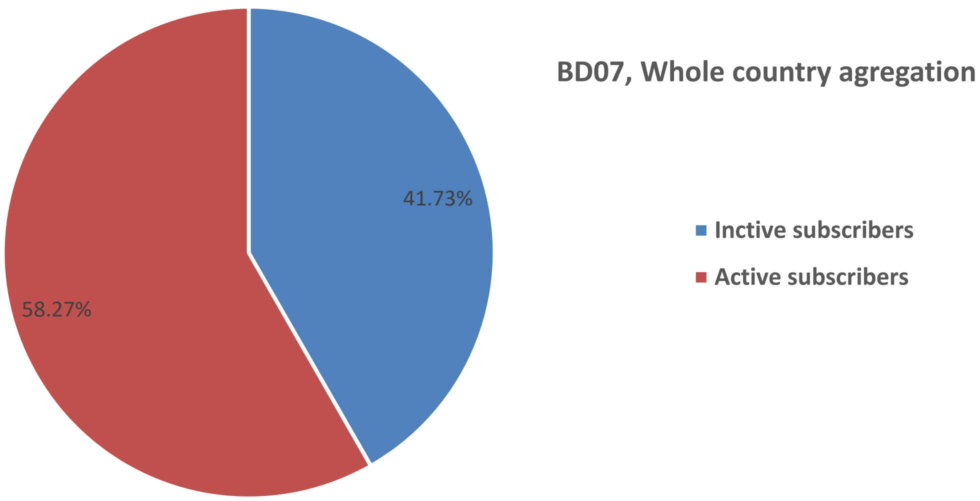
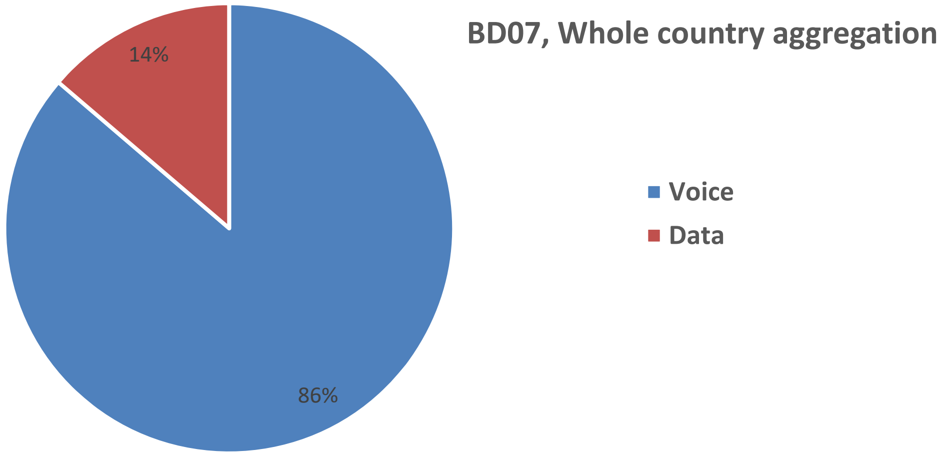


Figure 35: Average daily event type activity for the whole country of Georgia



5.8. 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 Equipment Identity (IMEI) code and respective Type Allocation Code.

In figure 36, the distribution of active mobile devices for the whole country of Georgia is provided. The most common devices across the country are handheld phones, i.e. old generation phones. Smartphones, however, are the second most used devices. As shown in figure 37, handheld phones are the most popular devices in all the regions except Mtskheta-Mtianeti. In figure 38, figure 39 and figure 40, the spatial distribution for, respectively, handheld phones, smartphones and tablets are provided. Such distributions are quite similar, meaning that there are no difference between the penetrations of handheld and smartphones, as already observed from LAU1 regions (figure 36). The total number of active devices in Georgia is approximately 5 million.

Figure 36: Active mobile devices for the whole country of Georgia

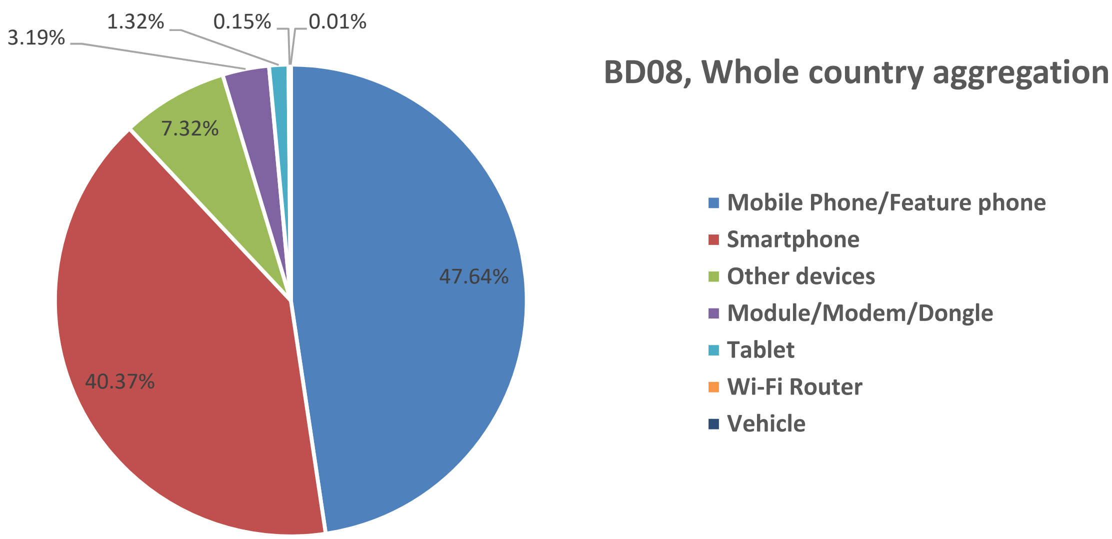


Figure 37: Number of active mobile devices by LAU1 regions

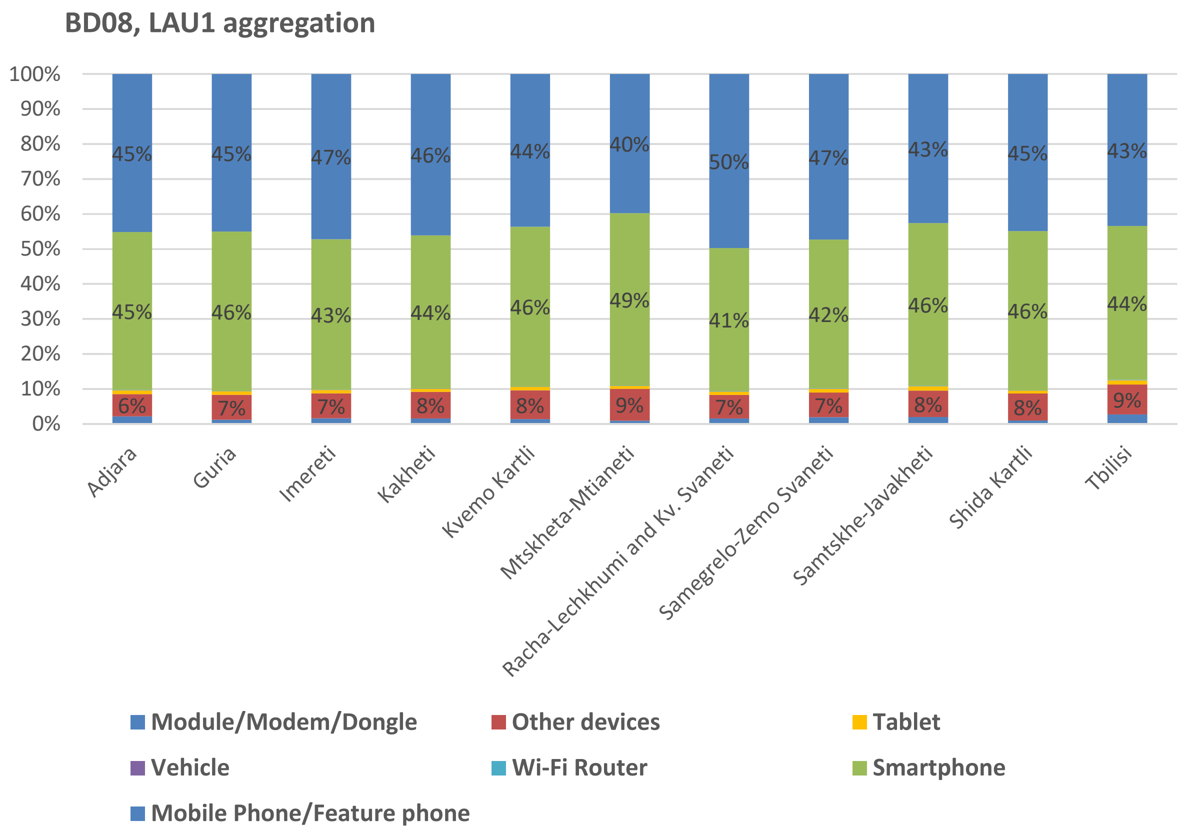


Figure 38: Number of active mobile handheld phones in LAU2 districts

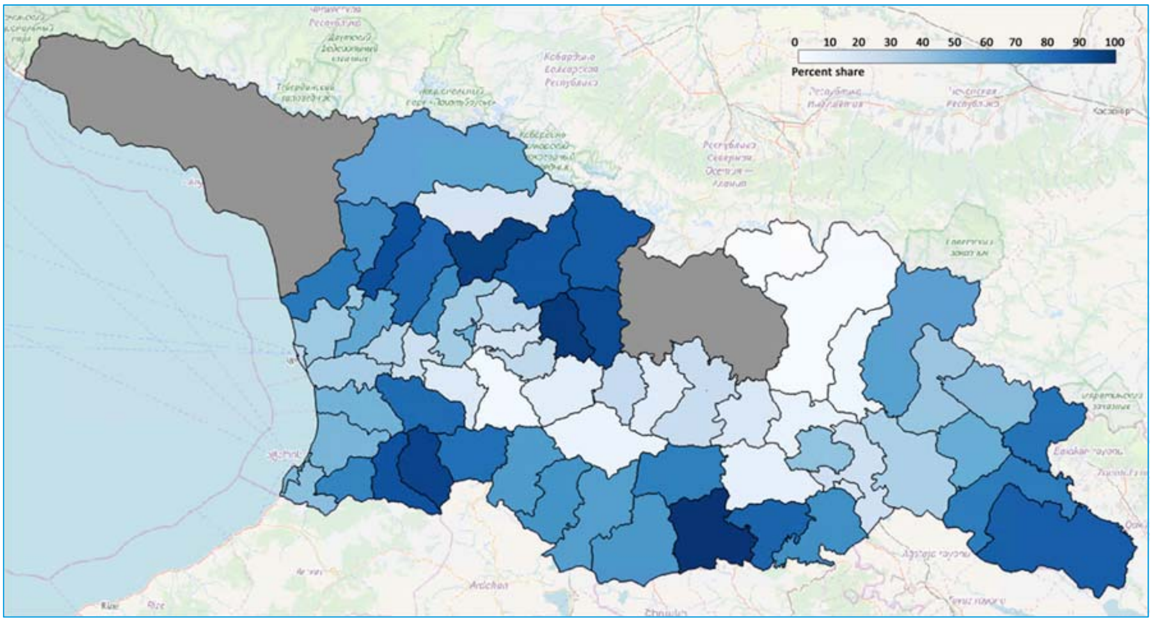


Figure 39: Number of active smartphone devices in LAU2 districts

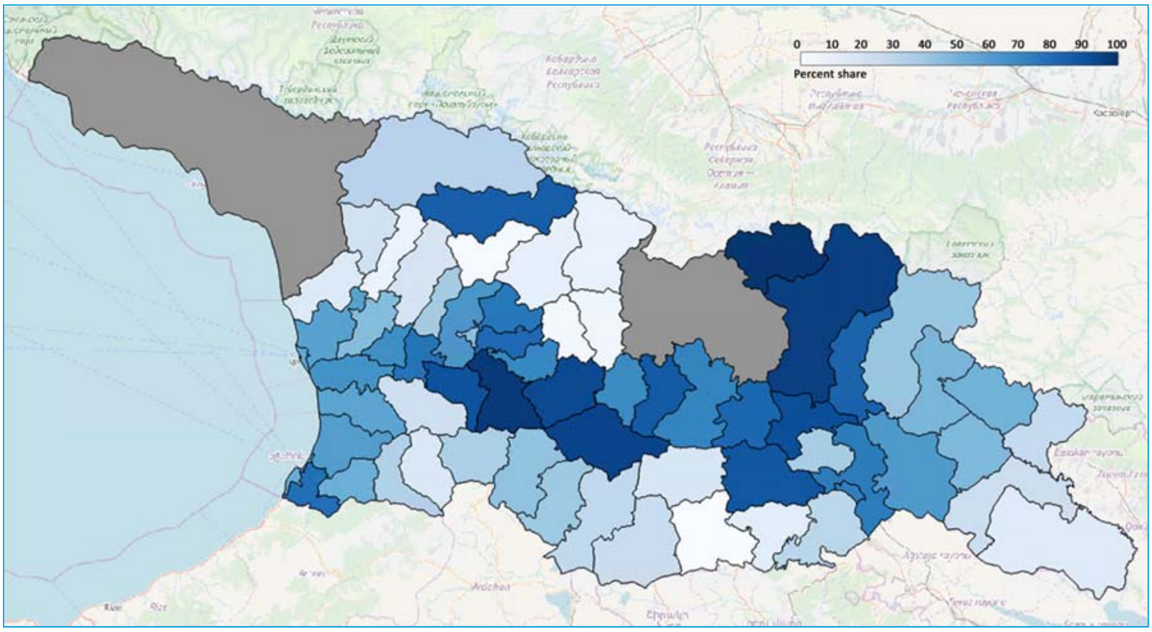
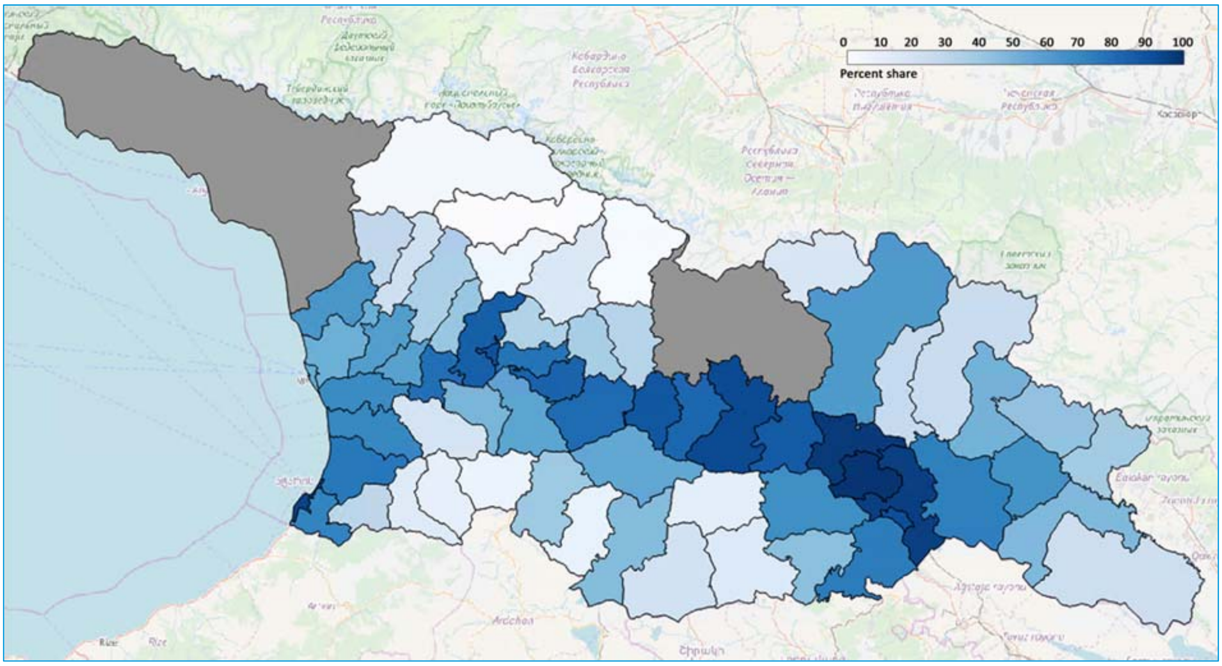


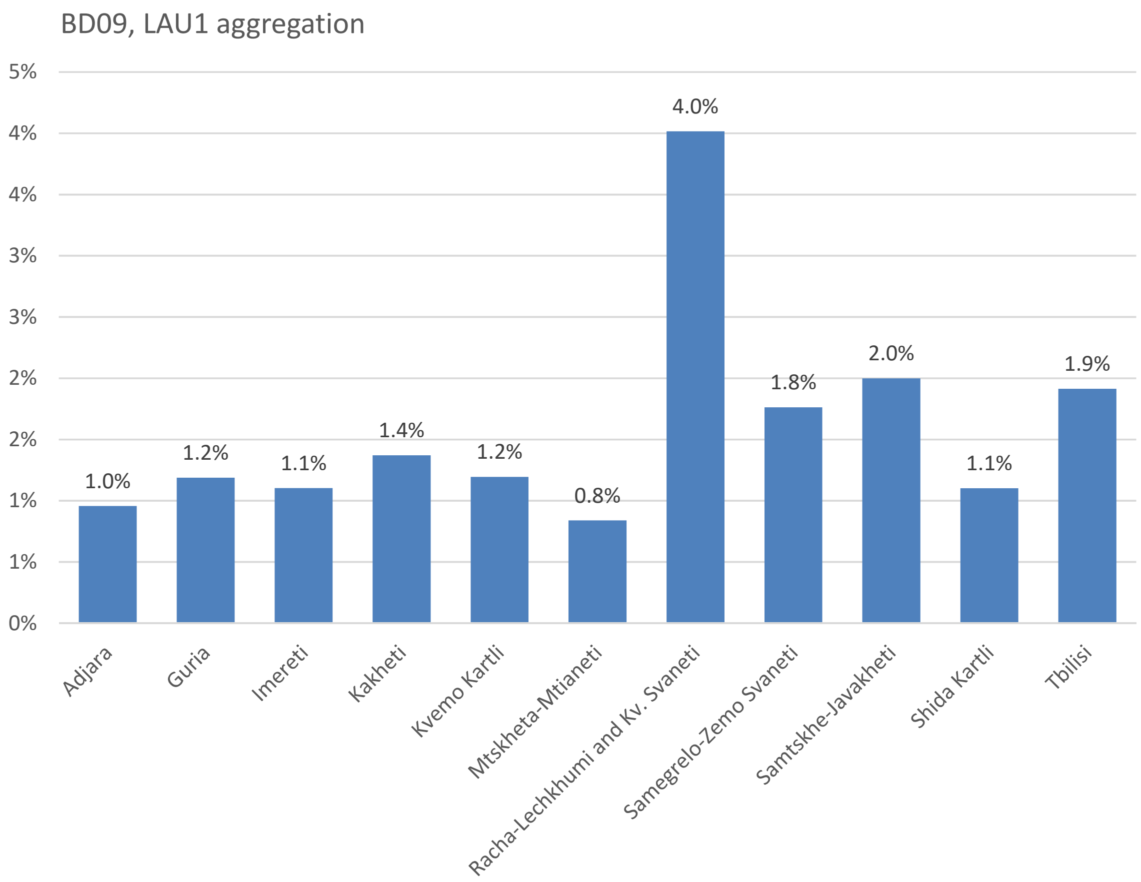
Figure 40: Number of active tablet devices in LAU2 districts



5.9. BD09: IMEI CONVERSION RATE

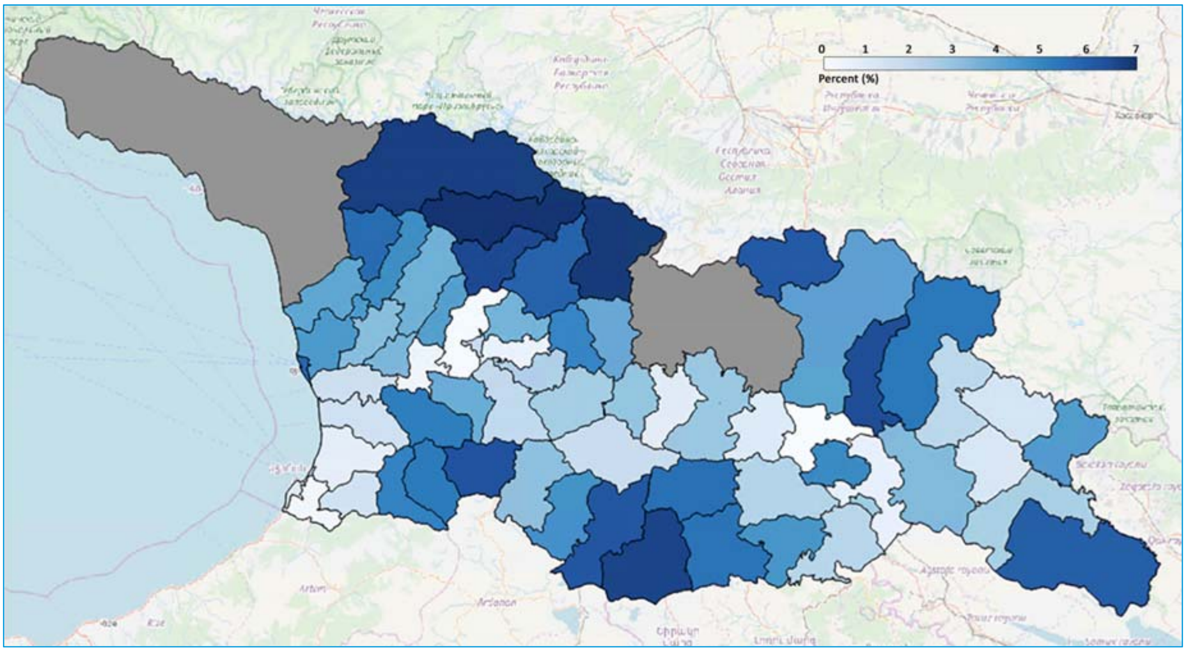
This is a new indicator describing the conversion of mobile devices in the country's mobile networks. Ideally, the IMEI code represents unique devices used in the network. Subscribers change the devices over a period of time, so this indicator should ideally represent the number of subscribers that have changed the 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 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% (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 41: IMEI conversion rate (how many subscribers changed the device) by LAU1 regions



From figure 41, it is clear how the percentage of users changing IMEI during the IV quarter of 2016 is below 2% for almost all the regions. The only exception is represented by Kv. Svaneti (4%). However, by aggregating with a finest spatial grain, i.e. LAU2 districts, some regions with percentages greater than 10% can be spotted (see figure 42). This indicator suggests that subscribers in Georgia do not change their mobile devices very often and probably use the bought device for a long time. This also suggests that there is no black market for “stolen” and regular changing of devices’ IMEI codes (to avoid service providers’ blacklists for stolen IMEI codes).

Figure 42: IMEI conversion rate (how many subscribers changed the device) by LAU2 districts



5.10. 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.

Overall, most of the traffic volume is generated by LTE networks in Tbilisi. This is rather obvious, since the higher bandwidth of LTE allows for higher traffic volume in the same amount of time. However, as depicted in figure 44, in most LAU1 regions outside of Tbilisi, the 3G network generates more traffic volumes compared to LTE networks. Tbilisi is the most populous region in Georgia, resulting in higher traffic generated. As highlighted in figure 45, figure 46 and figure 47, traffic volume generated in the Tbilisi region is higher than any other regions by a wide margin. The total traffic volume in Georgia is 5.9 petabytes.

Figure 43: Mobile domestic broadband traffic, by technology in petabytes, Whole country

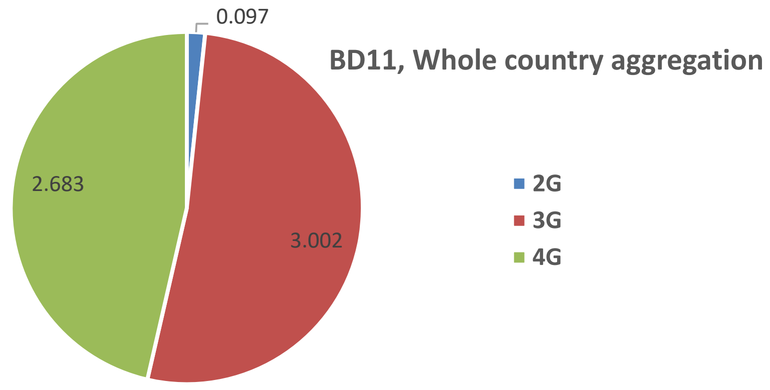


Figure 44: Mobile domestic broadband traffic, by technology, aggregated by LAU1 regions

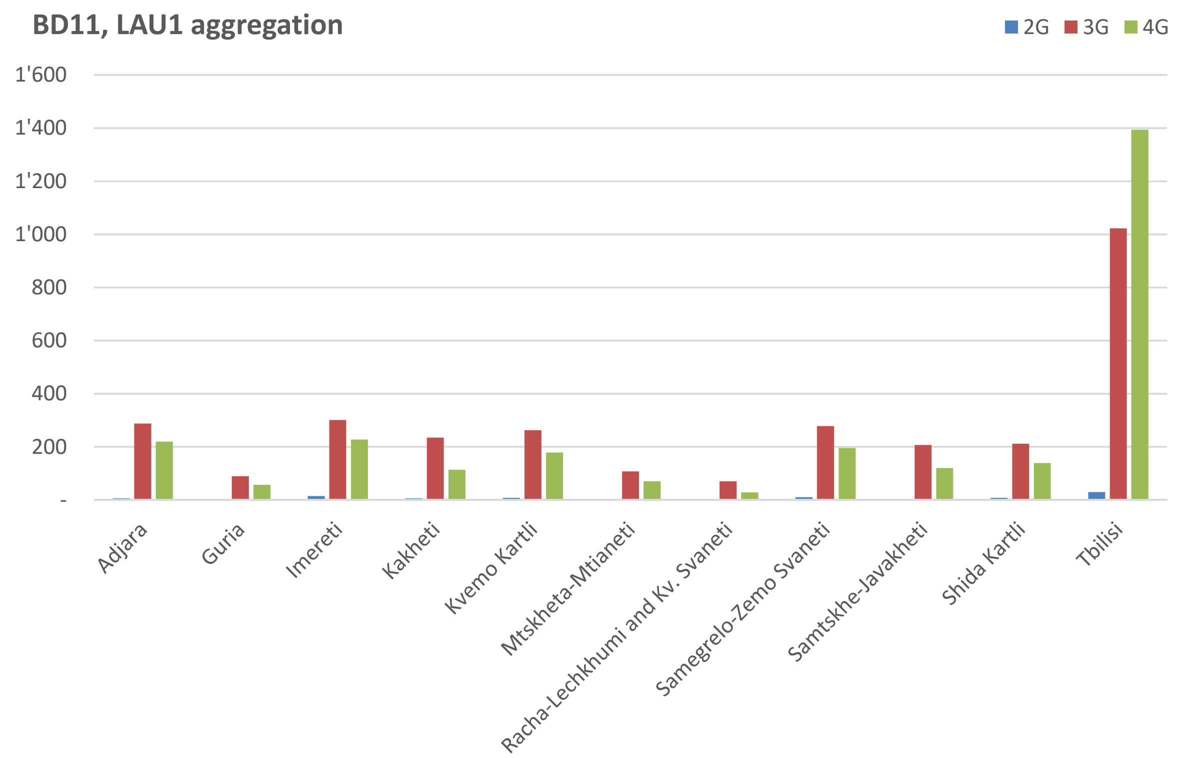


Figure 45: Mobile domestic broadband traffic for 2G network in LAU2 districts

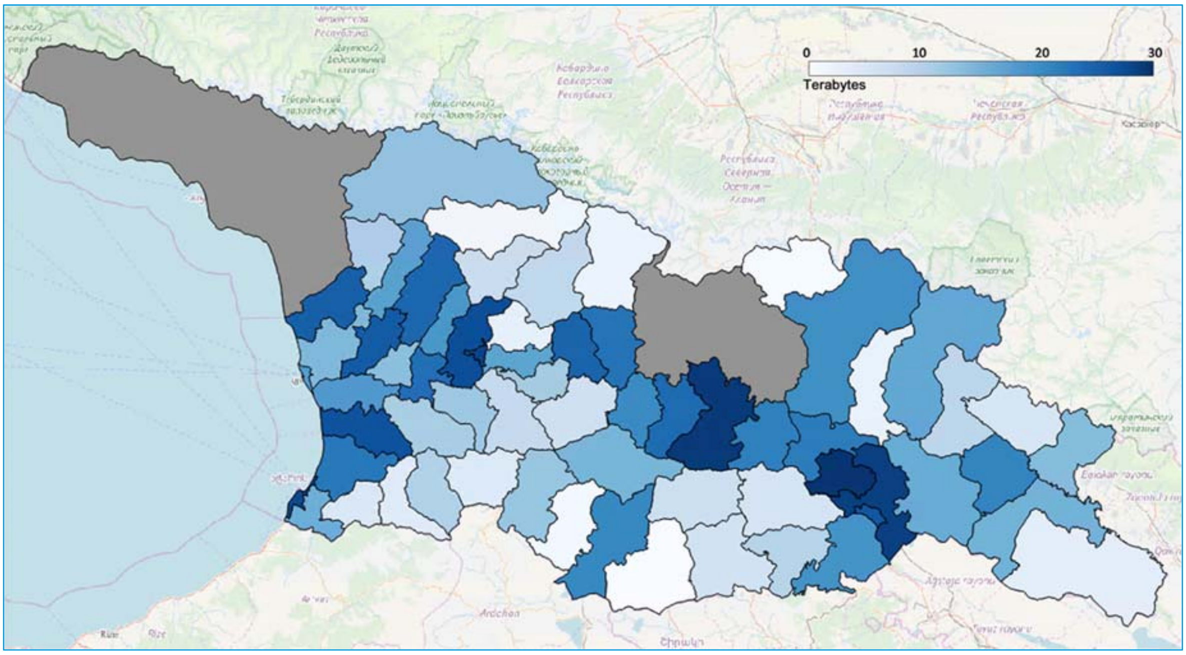


Figure 46: Mobile domestic broadband traffic for 3G network in LAU2 districts

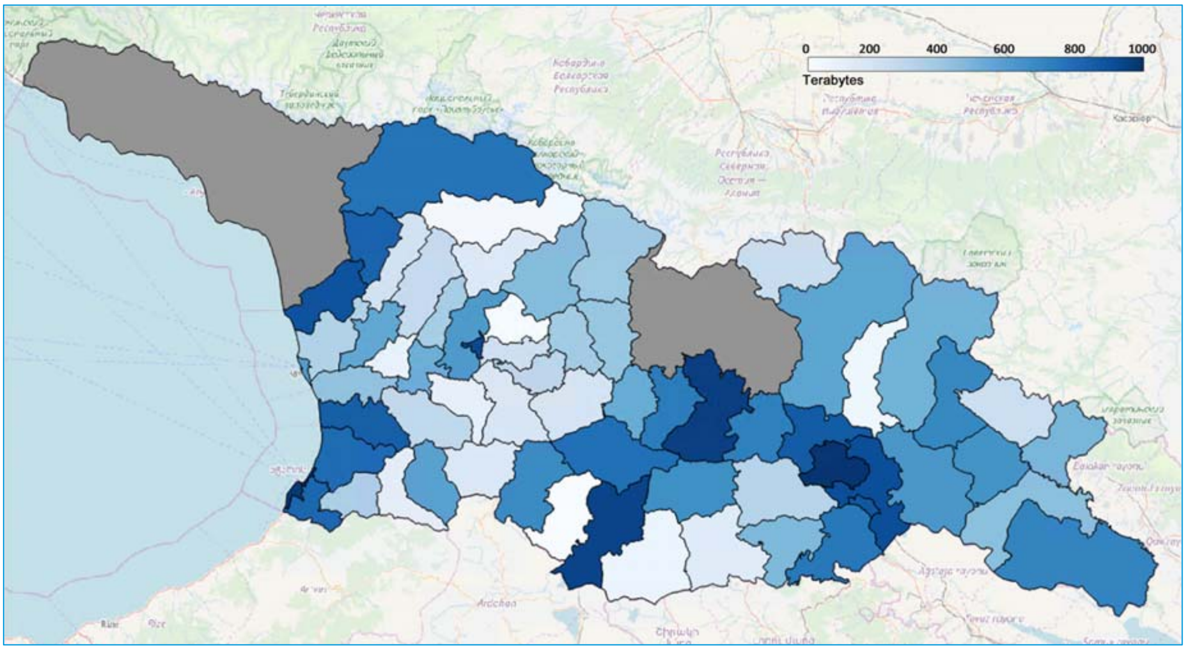
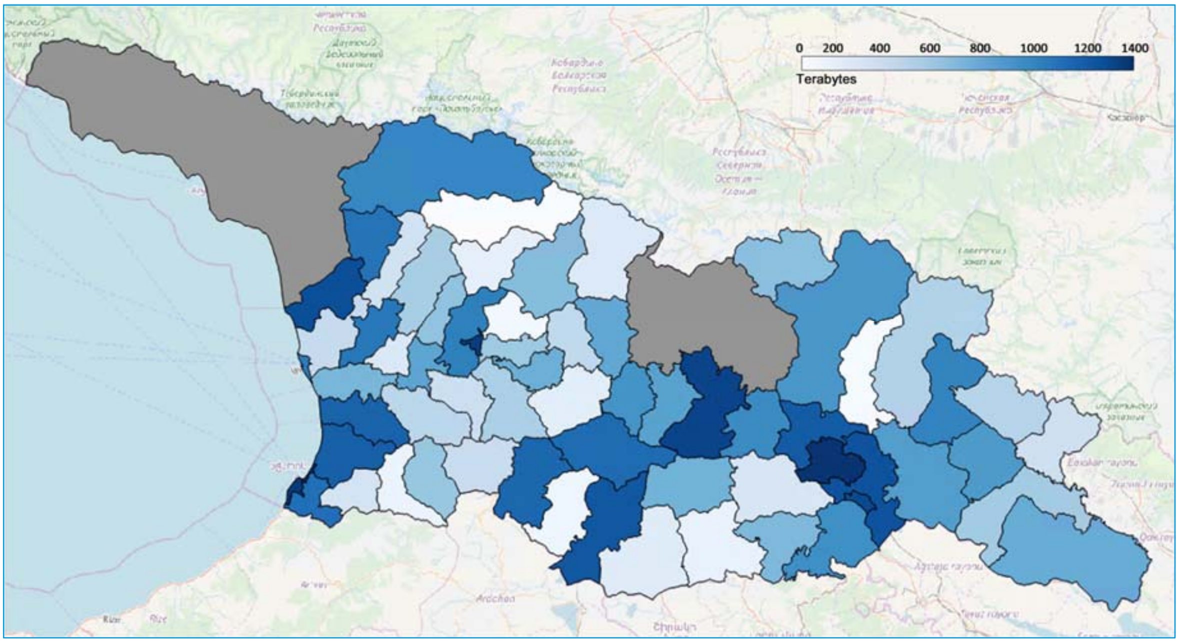


Figure 47: Mobile domestic broadband traffic for LTE network in LAU2 districts

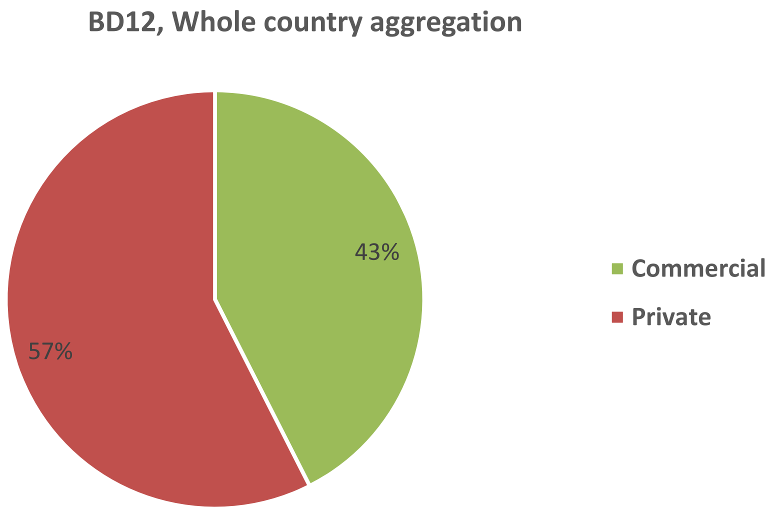


5.11. BD12: MOBILE INTERNATIONAL BROADBAND TRAFFIC, BY CONTRACT TYPE

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 operators roaming 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.

Results are showed in figure 48, for both commercial and private contracts. The whole broadband traffic used by Georgian subscribers abroad during the period is a bit over 1 petabyte. Subscribers with private contracts use about 57% and commercial subscribers 43% of this traffic volume.

Figure 48: Mobile international broadband Internet traffic for the whole country of Georgia



5.12. BD13: INBOUND ROAMING SUBSCRIPTIONS PER FOREIGN TOURIST

This is a new indicator representing the number of foreign roaming subscriptions compared to the number of foreign tourists accommodated in the country.

The number of foreign roaming subscriptions in the country correlates to the number of tourists visiting the country. Depending on the cost of the roaming service, the number of roaming subscriptions might be higher or lower – some foreigners tend to obtain a prepaid local SIM card for the duration of the stay, therefore not using a roaming service. The practice of reducing the costs of roaming services has resulted in a higher number of foreigners using roaming services, and this indicator (if used for a longer period of data) should reflect the effect of roaming service costs.

The indicator (division of roaming mobile subscribers to the foreign tourists) represents the estimation of how many tourists use roaming services in the country. Although it is not definitive that this is the exact proportion of tourists who use roaming, because the concrete tourists cannot be linked to concrete mobile subscriptions, tourists might have several mobile devices, no devices or use local SIM cards, and Georgians who have foreign SIM cards are not considered as tourists. Still, the majority of inbound roaming subscribers are assumed to be foreign tourists. Ideally, with decreasing roaming costs, more and more tourists do not buy local SIM cards and instead use roaming services, in which case ideally this indicator should be getting close to 1.0.

Figure 49: Incoming international travelers (tourists) and unique inbound roaming subscriptions for the whole country of Georgia, by months

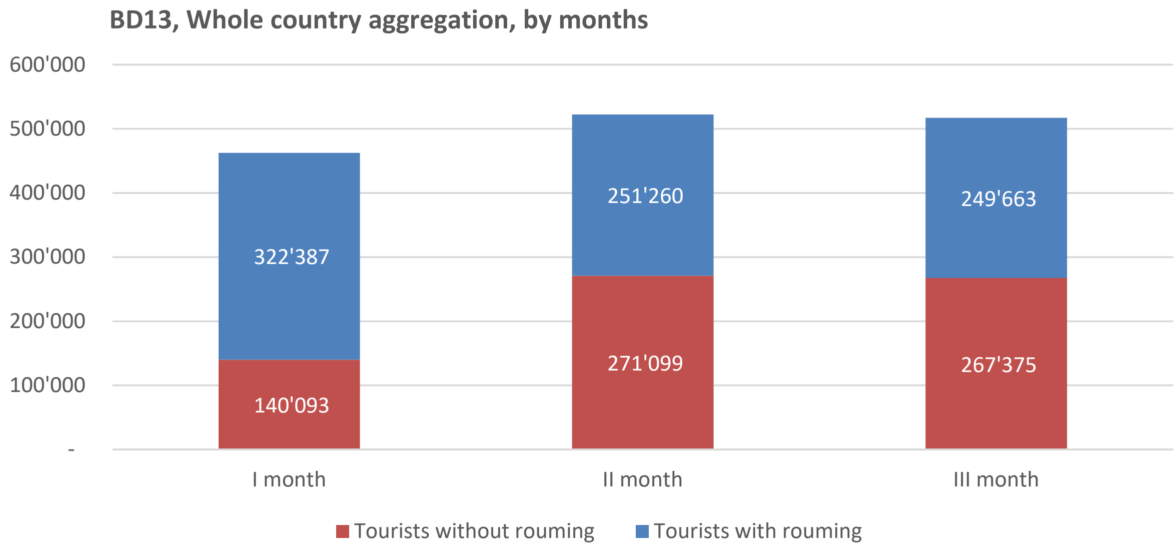
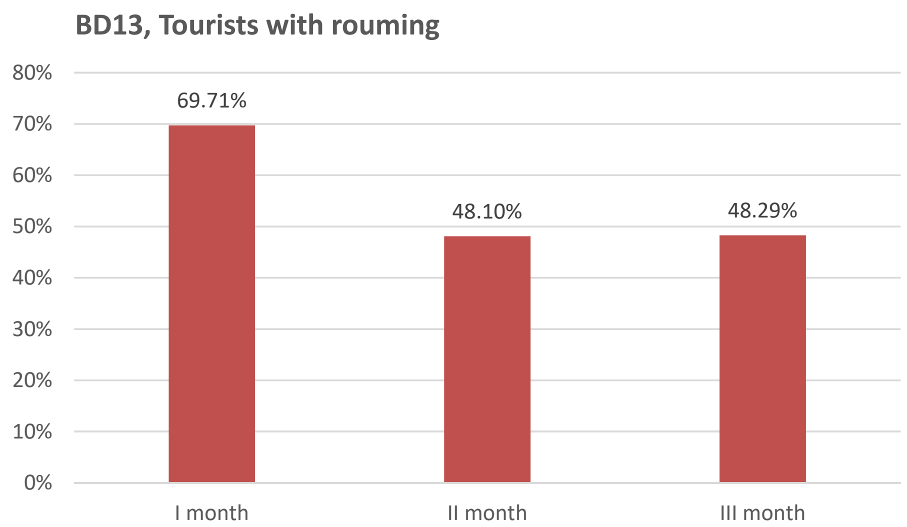


Figure 50: BD13 indicator representing the proportion of tourists using inbound roaming services in Georgia



The number of inbound roaming subscribers could also be used as input to official tourism statistics, making it possible to generate tourism statistics not only on a national, but also a regional scale, if this number could be generated regularly and with more granular spatial and temporal aggregation (month, week, day/region, cities and locations).

Figure 51: inbound roaming subscriptions for the whole country of Georgia, by visitor's country

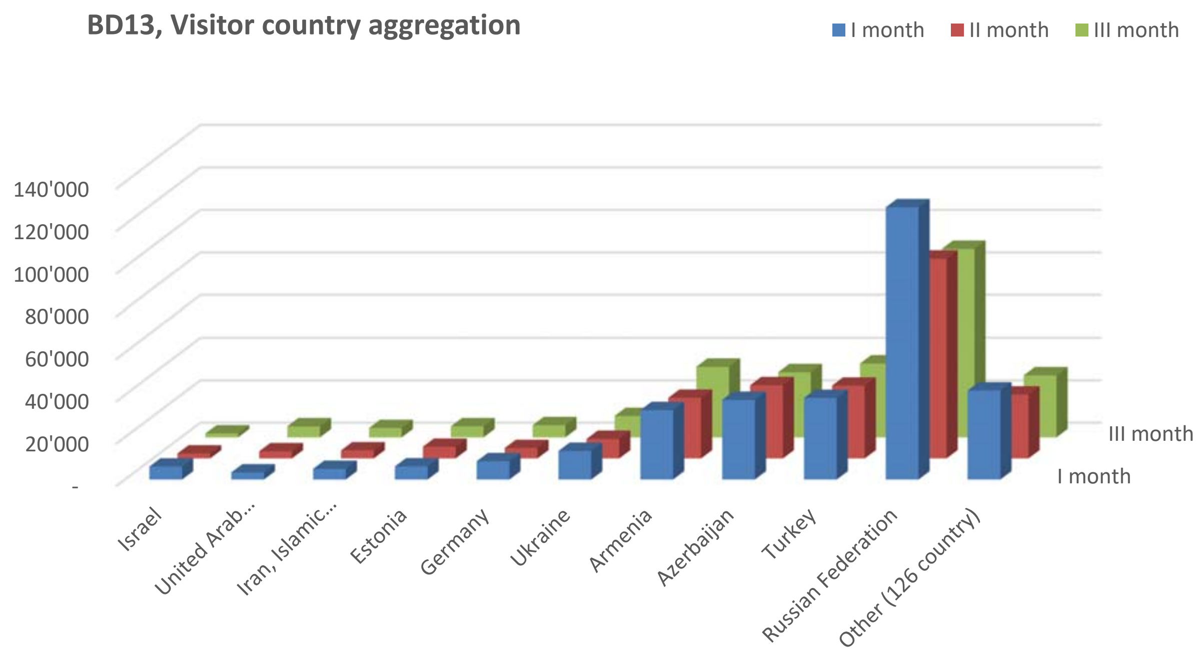
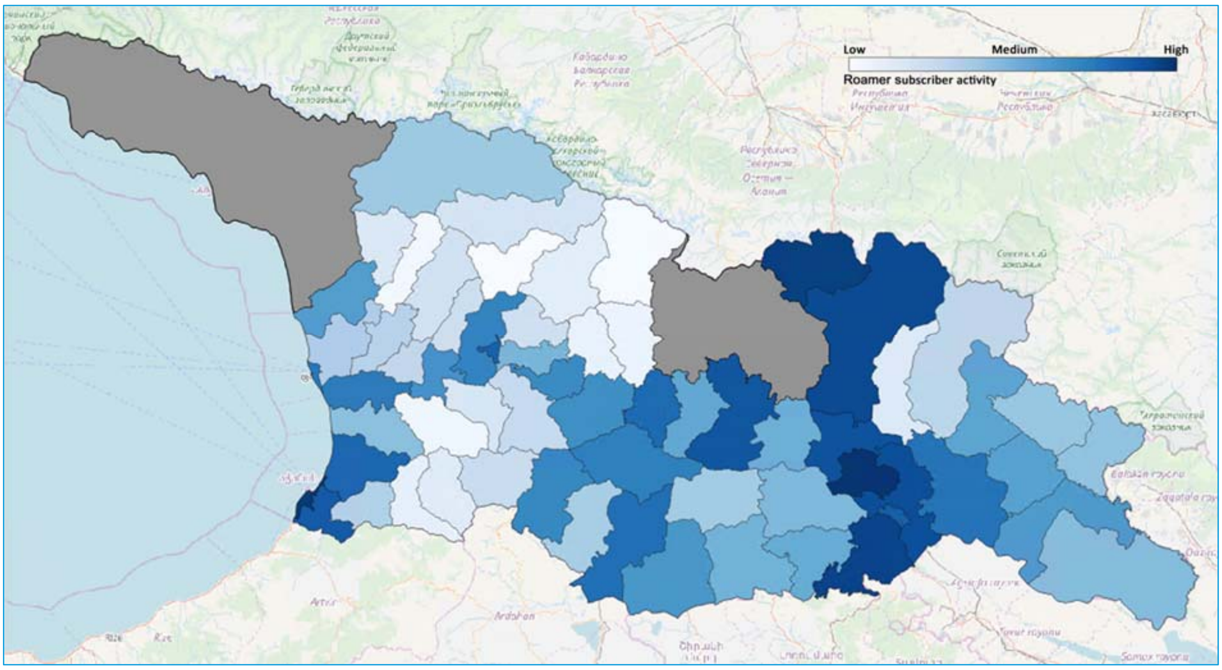


Figure 52: Inbound roaming subscriptions activity by districts



5.13. BD14: FIXED BROADBAND SUBSCRIPTIONS, BY TECHNOLOGY

These indicators represent fixed broadband subscriptions, aggregated by technology. The indicator can be broken down as follows:

- DSL Internet subscriptions (COPPER);
- Fiber-to-the-home/building Internet subscriptions;
- Wireless-broadband subscriptions (P2M & P2P Wi-Fi).

The total number of fixed broadband subscriptions is about 660 000. In figure 54Figure 53, the distribution of broadband subscriptions, aggregated by LAU1 regions, is provided. Fiber is the most used technology, with particular focus on the Tbilisi region, where the majority of subscribers are using a fiber connection. LAU2 disaggregation is provided in figure 55, figure 56 and figure 57.

Figure 53: Number of fixed broadband active subscriptions by technology, Whole country

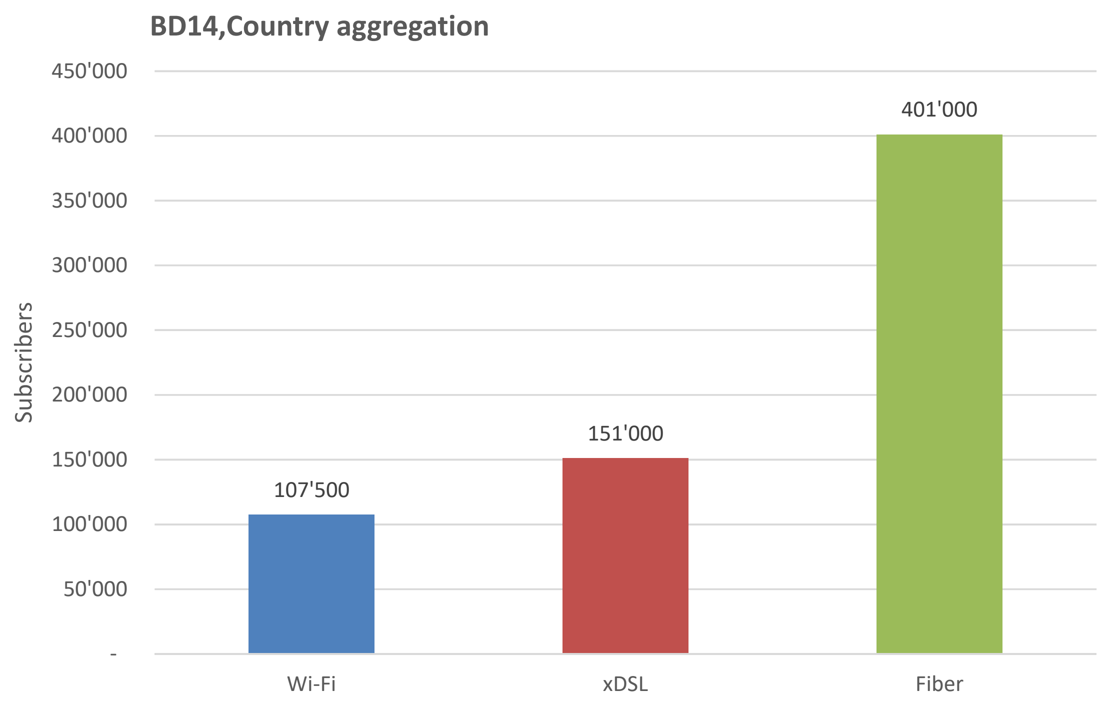


Figure 54: Number of fixed broadband active subscriptions in LAU1 regions

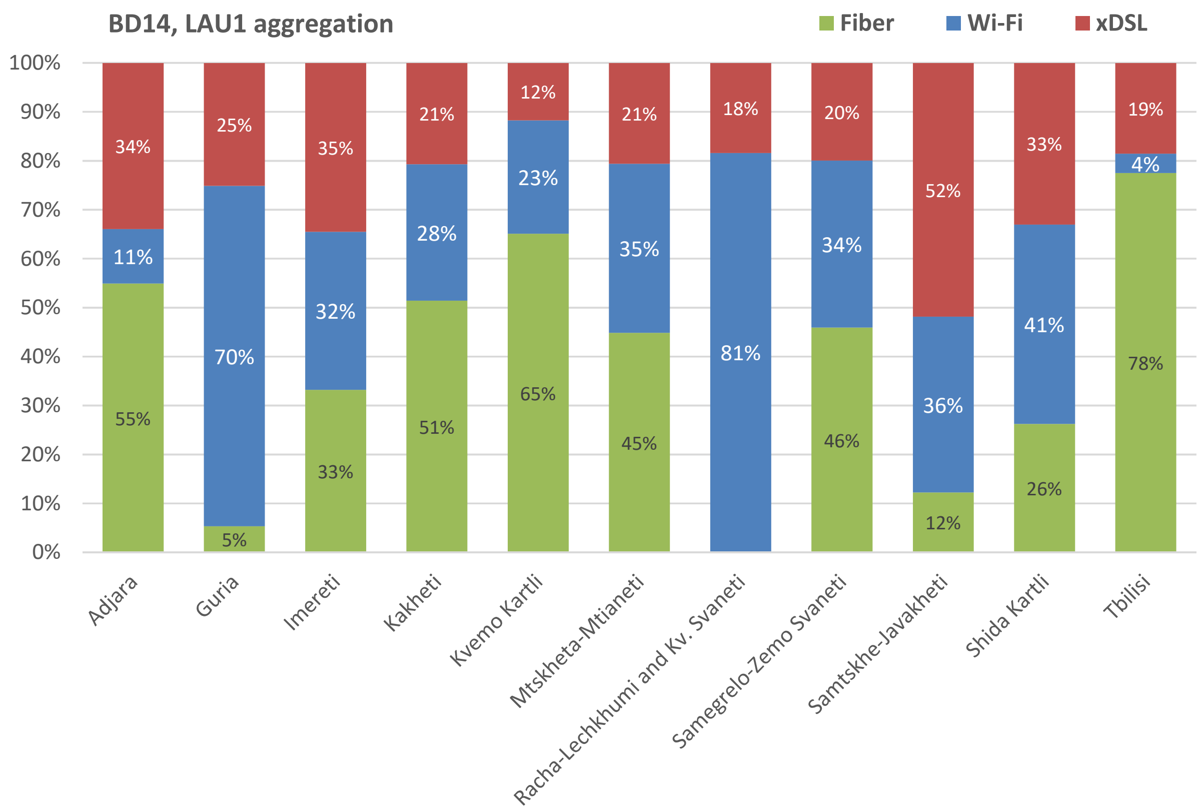


Figure 55: Broadband active subscriptions for xDSL technology in LAU2 districts

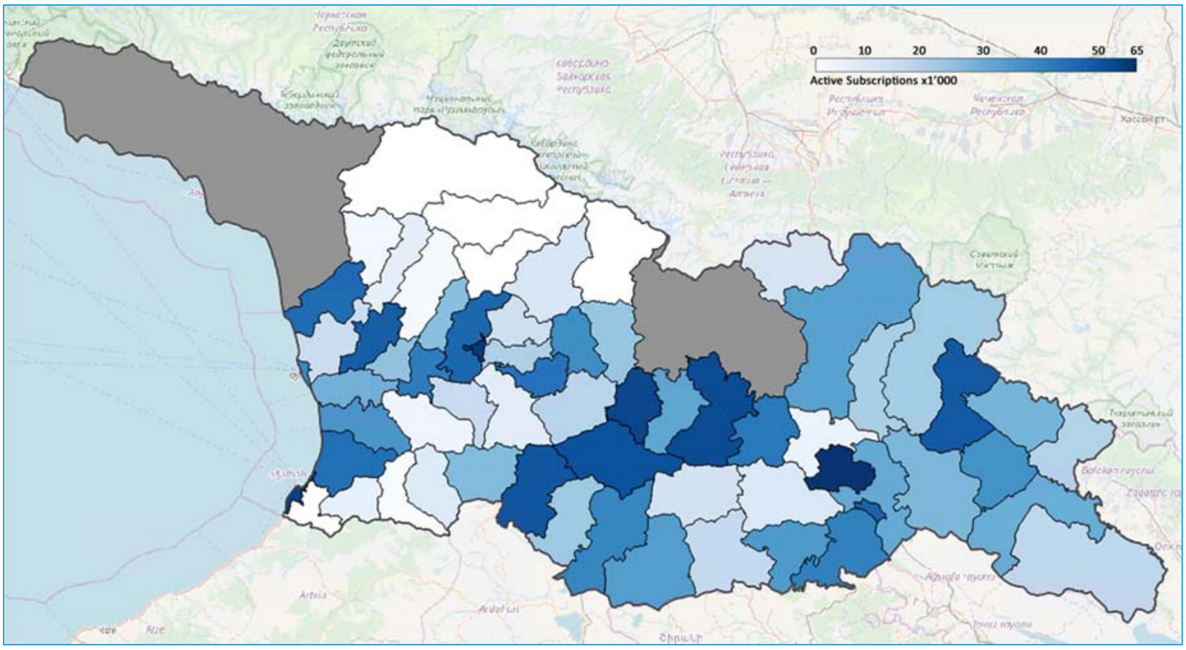


Figure 56: Broadband active subscriptions for fibre technology in LAU2 districts

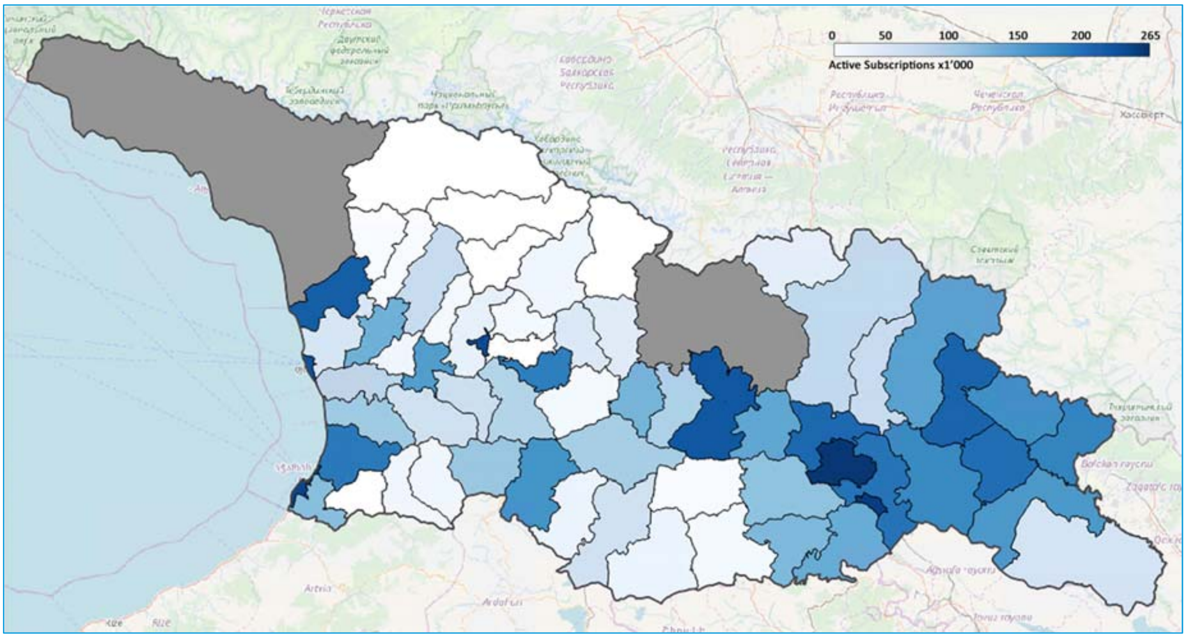
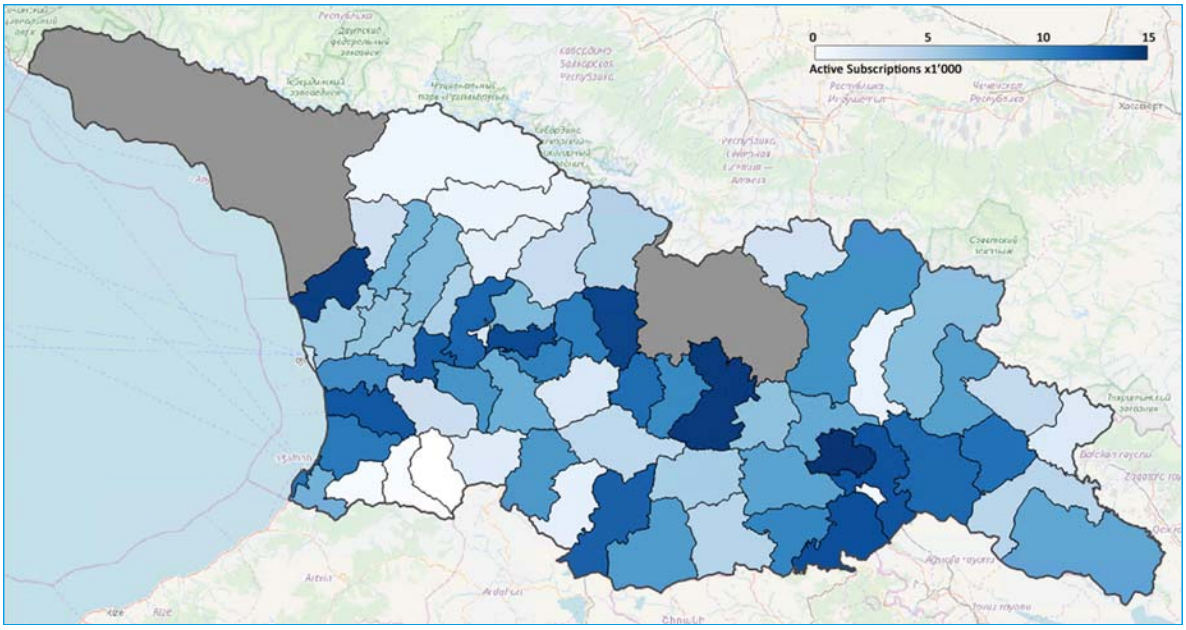


Figure 57: Broadband active subscriptions for fixed Wi-Fi technology in LAU2 districts



5.14. 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.

As shown in figure 59, the highest speed, 30–100 Mbit/s, is only available for subscription in the Tbilisi and Adjara region. Other regions had subscriptions with a maximum speed of 10 Mbit/s. Since 2-10 Mbit/s is the most common subscribed speed, in figure 60, figure 61 and figure 62, the spatial distribution for such speed is provided. Data are aggregated by LAU2 districts, showing how Tbilisi, Batumi and Rustavi are the regions with the most subscribers for this type of connection speed.

Figure 58: Fixed broadband subscriptions, Whole country

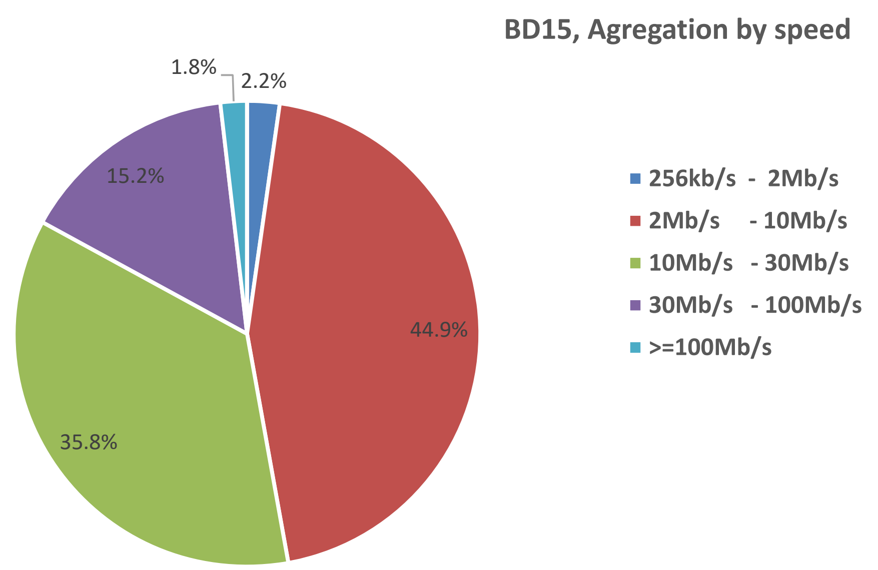


Figure 59: Fixed broadband subscriptions, aggregated by speed and LAU1 regions

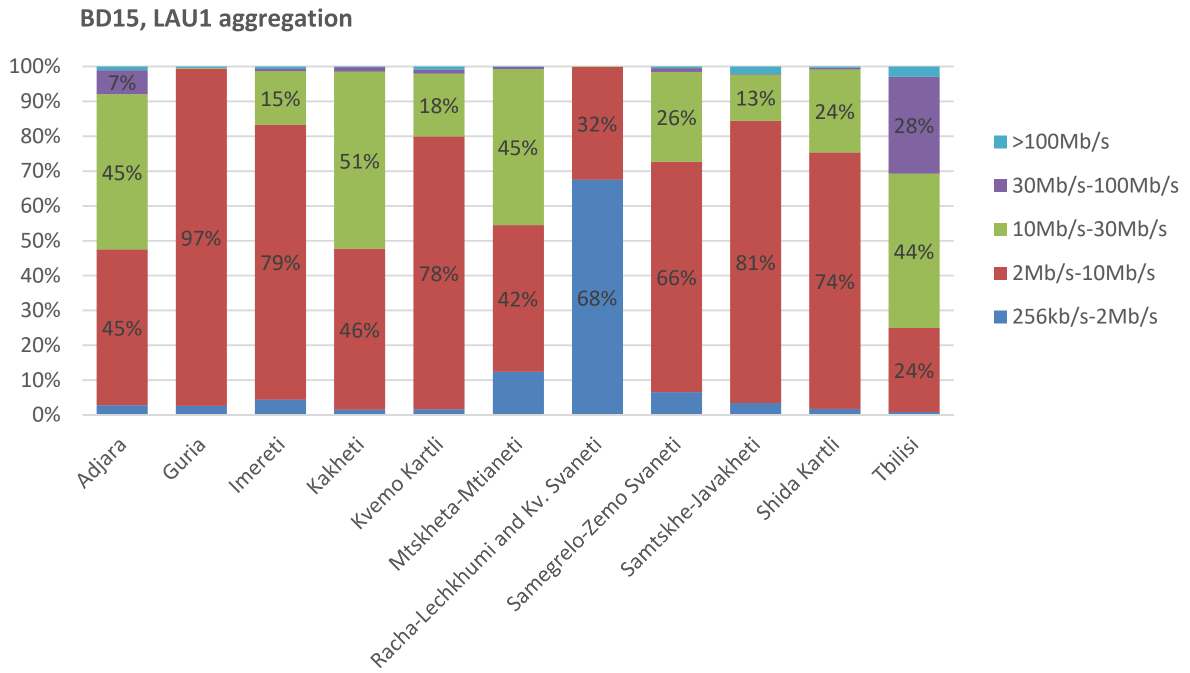


Figure 60: Fixed broadband active subscriptions for 2–10 Mbit/s speed in LAU2 districts

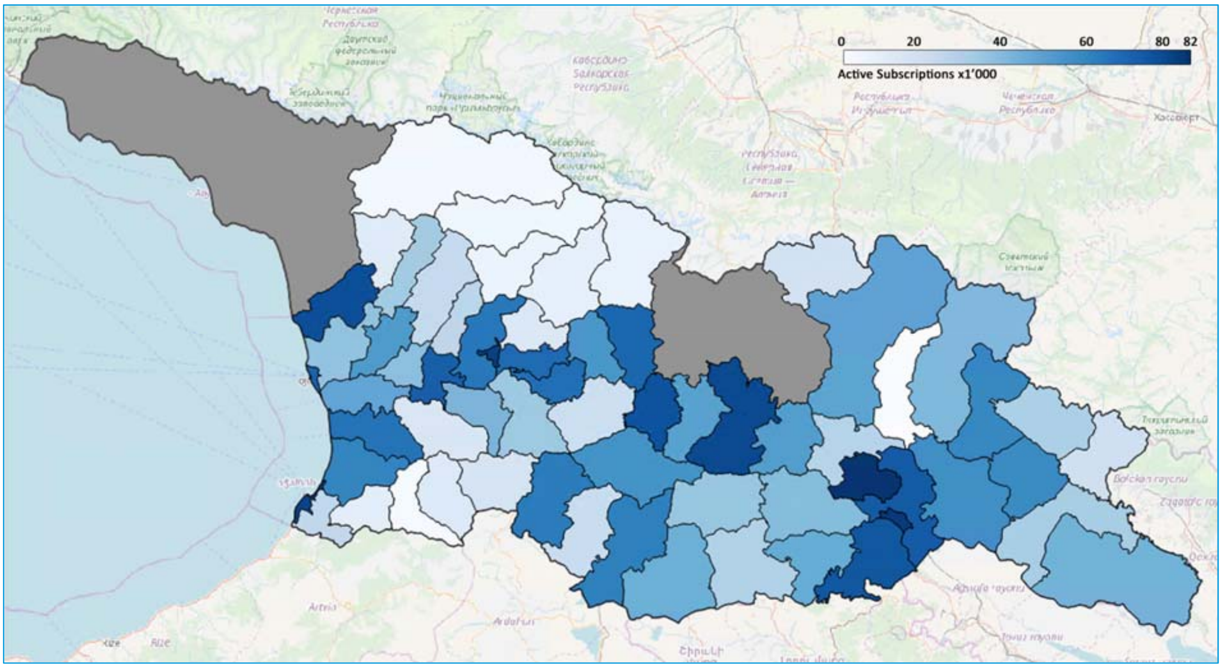


Figure 61: Fixed broadband active subscriptions for 10–30 Mbit/s speed in LAU2 districts

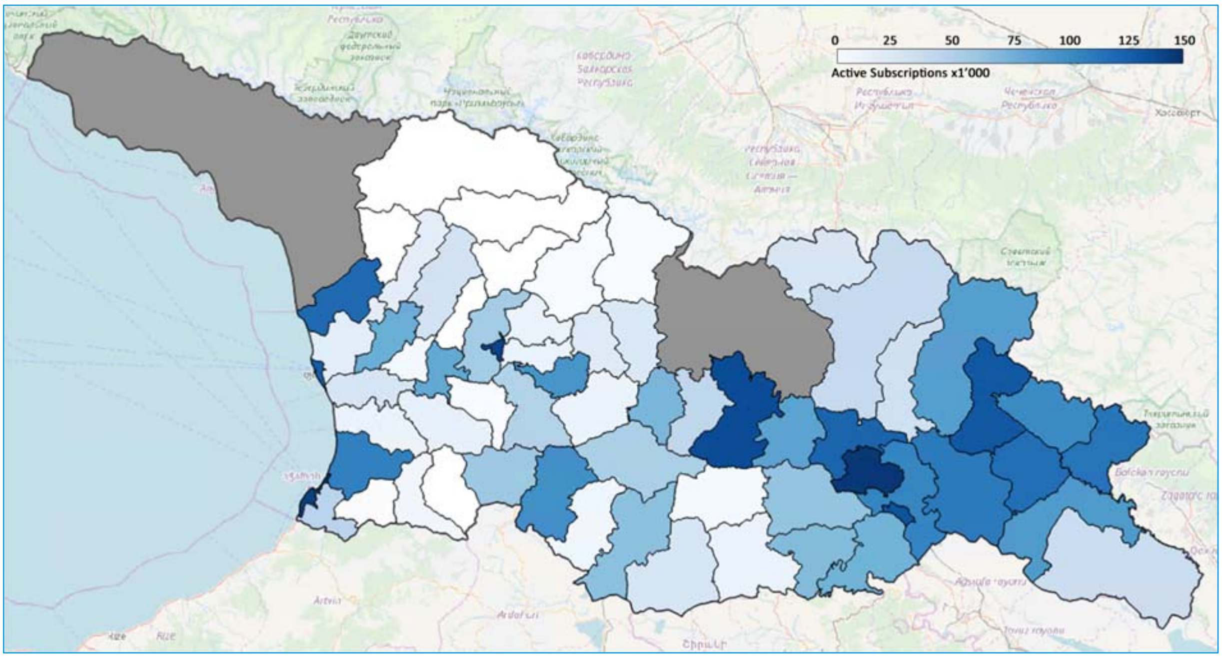
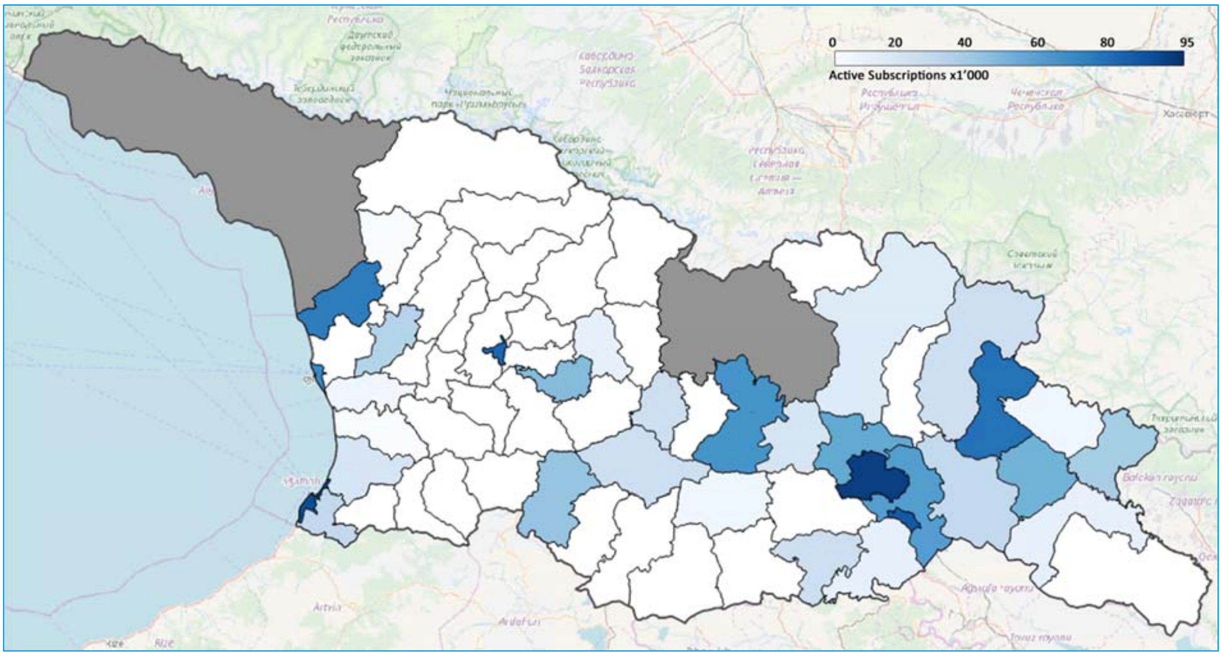


Figure 62: Fixed broadband active subscriptions for 30–100 Mbit/s speed in LAU2 districts



6. RECOMMENDATIONS

Based on the experience of the pilot project in Georgia, 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 (memoranda of understanding, non-disclosure agreements, 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 data protection authority's (or other privacy oversight organization's) involvement in the project is important as well to make sure the processing of the data is conducted according to the legislation.
- The specification, sources and business glossary of the raw data have to be specified and agreed upon between all data providers and stakeholders, including exact naming of the elements (e.g. what does CDR represent exactly, what attributes are used in each data record, etc.).
- The methodology of the processing of the data for each indicator needs to be defined and confirmed by all stakeholders well in advance. The general understanding of how the indicators will be processed, what are the intermediate tables, how should they be aggregated, what are the breakdown dimensions, etc. – these questions need resolving before the actual processing takes place. This often requires a pilot project or iteration of methodology development before the final methodology can be agreed and locked for the "real" indicators.

- 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 the 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).
- Algorithms (even if they are prepared as examples) used to calculate the indicators have to be prepared along with the methodology, so no difference or misinterpretation of how the indicators should be calculated can appear. Again, this might require a pilot project for adjusting the methodology and the algorithms several times, but the interpretation of the methodology by different telecommunication service providers might cause serious discrepancies in resulting indicators.
- In order to produce valuable and comprehensive analysis and relevant reports it is necessary, moreover vital for all mobile operators to use the same data depersonalization methodology. This methodology shall be valid few years period in order to be able to analyze the seasonal, geographical and other important trends.
- 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; however, 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 problematical if there is limited information on the technical algorithms used for calculating these indicators. In case validation of the indicators is necessary it is important that national statistics office is involved.
- 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 data scientist was able to visit the pilot country only once. This type of project would require multiple visits with intermediate calculations of the indicators, assessing the results and adjusting the methodology and algorithms, moreover it will be better if GNCC would have own Big-data data scientist, in order for, data processing to be faster and more productive.
- Agreeing on the source data, methodology, specific algorithms, business model and validation procedures take a lot of time. Thus, it is reasonable to plan an intensive preparation period including several meetings between stakeholders' business, 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.

- In case the data provided by the telecommunication operators envisages any information which is identified/approved as business secret, agreeing on the scope of the publication of the resulting indicators in terms of confidentiality of the telecommunication service providers is important to ensure the safety of the business secrets of the telecommunication service providers.
- 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 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 an initial assessment from the telecommunication service providers 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; however, further 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.
- Worth noting, that the fact that GNCC collected the Tier I data and calculated the indicators themselves was very time savvy and gave the possibility to adjust and recalculate the indicators where errors occurred in timely and accurate manner. Therefore, it would be very important to always have this possibility for a party to calculate the indicators themselves.
- The value of the big data does lie not simply in the production of indicators with more breakdowns, it also provides 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 behavior changes in the consumption of ICT could provide insights into the trends and foresights (predictions) of the use of ICT – not simply by looking at the overall trends of usage, but clustering and classification of the users based on their individual behavior 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. 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 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 about how use of big data can benefit 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 LTE and higher technology would increase to 20% (BD05), for the whole country in the next year, and in no administrative subdivisions this number should be below

12%. The specific target objectives are obviously country-specific and can be agreed between local stakeholders.

- During this project GNCC has used relational databases while working on Big-data. In future it is highly recommended to use non-relational NoSQL technology databases, because working with big-data is more productive when using special databases which are recommended for Big-data.
- Worth noting, that reports from Big-Data are indeed complex and rich with variations and to draw such charts and maps in static reports has a lot of difficulties. Making it online, web based dynamic reports, charts and maps are more user-friendly.

7. CONCLUSIONS

The pilot project in Georgia was able to produce majority of the proposed indicators in the methodology document. There were useful lessons generated relating to both procedural matters and statistical outcomes. These learnings were derived in the process of overcoming various challenges faced in project implementation. Such learnings are especially valuable to countries who wish to replicate and use mobile phone big data to derive official statistics.

In general, the challenges can be divided into:

- Administrative and legal; and
- Technical and methodological.

The administrative and legal challenges were successfully dealt with because of the strong previous experience of GNCC collecting similar types of data from 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 14 indicators calculated in this project give a good overview of the possibilities of the big data for use in ICT statistics calculation.

It must be noted that some indicators were not a match to officially reported indicators (e.g. BD02 and BD06), and the reasons for that might be in different methodologies how the official number is calculated.

For BD02, the difference can be caused by the method of calculation – Georgia is a country covered by many mountains which may be sparsely populated, so perhaps the almost perfect coverage reported by official indicator (91–99%) could include only the land where people live and exclude the less populated mountain regions.

For BD06, this project included all subscriptions (voice and data), and the data were received from all MNOs. However, the data represented only those subscriptions that actively used network service (either voice or data). Some active subscriptions in official statistics might not have been conducting calls nor accessing the Internet, and thus were not accounted for in this project.

Indicators BD03 and BD04 show the proportion of usage of different mobile-cellular technologies for voice and Internet access. Comparing those indicators to BD05 shows a large difference of usage and access.

The main conclusions from those indicators suggest that those who have access to 3G and LTE networks (≈ 40%) use those technologies actively to access the Internet (BD04), but there are still very many who for some reason do not have access to LTE due to coverage issues or their contract limitations.

Roughly a third of the subscribers are active on a daily basis. The reason for this is unknown, and ideally, all subscribers should be active on all days.

Georgian subscribers do not change their mobile devices very often. Only 2% (BD08) have changed their mobile devices during the observation period, and this suggests Georgian subscribers do not buy new phones very often.

The indicators related to Tbilisi are in most cases higher than in other regions of Georgia, which is expected, given that Tbilisi is the most urbanized part of Georgia and the main business hub. Regional differences can be observed for all parts of Georgia, and this can be used to direct investments in ICT to less-advanced parts of the country.

The big data pilot project has demonstrated that the data obtained from the mobile CRD/IPDR, in addition to the processed indicators, are rich with many attributes which are both interesting and valuable for telecommunications as well as non-telecommunications purposes. The success of the pilot project could be attributed to the strong experience of GNCC in collecting the data, as well as its work on ICT indicators. Accordingly, if a similar exercise will become a permanent area of work of the GNCC, it can reveal many different and valuable statistics and analytics.