



Use of mobile devices and other advances in data collection

Tuulikki Sillajõe*

Statistics Estonia, Tallinn, Estonia – tuulikki.sillajoe@stat.ee

Jaanus Kroon

Eesti Pank, Tallinn, Estonia – jaanus.kroon@eestipank.ee

Abstract

The paper describes the aspirations of Estonian official statisticians towards using new data sources.

The paper also provides a short overview on the results of an international consortium's (which Statistics Estonia was participating in) study, which focused on the usage of mobile positioning data for tourism statistics. The strengths and weaknesses as well as the new possible uses of mobile positioning data are presented.

A concrete business case is described. Namely, in 2008, Eesti Pank (central bank of Estonia) started to develop a new data collection methodology jointly with the researchers of the University of Tartu, who were experienced in using mobile positioning data in urban and regional planning. After a two-year experience of using the data for external sector statistics, the methodology was considered to offer a reliable overview on persons crossing the Estonian border to travel abroad (outbound travel) and to Estonia (inbound travel). So, the official cross-border travel statistics have been published by Eesti Pank since 2012.

Keywords: big data; cooperation of official statistics, central banks and universities; mobile positioning; new data sources.

1. Introduction

Rapidly growing worldwide travels, membership in the Schengen Area, where no regular border controls and respective data collection are used between the member countries, and the termination of regular border surveys by Statistics Estonia due to budget cuts (the Quarterly Border Survey was taken out of the statistical programme as of 2010) have forced Eesti Pank, responsible for external sector statistics, to adopt an alternative way to continue the suddenly broken border-crossing time series. The latter data is an important input for the compilation process of the country's monthly/quarterly balance of payments (BoP) where the exports and imports of travel services play a remarkable role.

Various options (incl. road sensors, credit card information, accommodation statistics-based data models, etc.) were considered. To continue the time series, mobile positioning as the simplest and a relatively low-cost statistics instrument was opted for.

The choice was largely determined by the availability of a potential partner, the Department of Geography at the University of Tartu (Estonia), the spin-off company of which, Positium LBS, has continuous experience of using mobile positioning data in urban and regional geography/planning. Relying on the already established regular data exchange with the biggest mobile operator in Estonia and on the availability of the related calibration surveys, the scale-effect was expected to be remarkable. In 2008 Eesti Pank set up cooperation ties with the researchers at the University of Tartu to develop a new data collection methodology and models.

From the statistical point of view, mobile owners are a representative sample whose spatial behaviour and characteristics in time can be expanded to the total population. Theoretically, the data are readily available, and this makes data collection faster and more cost-effective. Because cell phones are widespread, the resulting data set is comprehensive, minimizing the human factor (interviewer's interpretation in surveys) and ensuring homogeneity, which improves the accuracy and quality of the data compared to traditional data collection methods.



2. Mobile positioning-based border-crossing statistics as a data source for BoP statistics

The jointly developed methodology is based on the use of the readily available log files of Mobile Network Operators (MNOs), which register the information necessary for the billing of roaming services (voice calls, SMSes/MMses, mobile data usage, mobile supported GPS, etc.).

In compliance with the BoP methodology, mobile positioning determines the residence of a traveller by the permanent residence criterion, regardless of the resident's citizenship or nationality. As a rule, entering into a contract with an MNO is the most favourable in the country where the phone is used the most frequently; therefore, the presumable residence of phone owners is determined by the registration country of the SIM card. This approach is supposed to give even more precise results according to the residency concept.

The amount, length and nature of the visits of Estonian residents and non-residents are determined by the location-based anonymised patterns of use of mobile phones according to the roaming activities in reporting resident operator network and operator clients' roaming activities in networks abroad.

Data processing broadly consists of the following, sometimes country-specific steps:

- Quality control of the data collected from the operator's system. Since the volume of data is huge, filters for finding and correcting errors were developed based on the characteristics of the data.
- Filtering and evaluation of the roaming data in order to ensure that the data are representative and of scientific quality.
- Geographical and temporal interpolation, i.e. linking additional parameters to ensure administrative and chronological comparability.
- Elimination of border bias (the crew of ships (seafarers) and transit (airport, transit routes).
- Profiling and segmenting of individual trips: single- and multi-day visitors; number of visits to and from the country, the country of origin (inbound) and destination (outbound) and the number of nights and days spent are calculated.
- Expanding to the total population by a special penetration model which takes into account the penetration rate and market share of mobile operators within roaming service, the results of calibration surveys on the differences of phone usage between residents from different countries, seasonality, etc.

According to the outsourcing contract, aggregated and referenced data for the previous month are provided to Eesti Pank on a monthly basis. The monthly results are verified and validated by Eesti Pank, taking also into account other official data sources (the number of passengers in the Port of Tallinn and at Tallinn Airport, crossings of the Estonian/EU administrative border, official accommodation statistics, the press, etc.). The data are used as an input for monthly and quarterly external sector statistics and have been disseminated as official statistics on Eesti Pank's website on a quarterly basis since 2012. The data are also supplemented with a link to the relevant press release.

The production of border-crossing statistics as a statistical activity is represented on the list of Eesti Pank statistical activities, a part of the statistical programme of Estonia. Comparing the current costs with earlier expenditures on regular border-crossing surveys with necessary interviews for producing similar statistics, the current approach is remarkably more cost-effective.

Although the collection of data on tourism is not a traditional task of the central bank, and the latter has no ambitions to go beyond the boundaries of its competence, Statistics Estonia is encouraged, on the basis of current experience, to take advantage of mobile positioning as an important data source in official statistics, especially in international and domestic surveys on tourism and the movement of people. In Estonia there are no legal constraints on considering MNOs as direct providers of statistical data regarding mobile positioning information for official statistics.

3. Feasibility study on the use of mobile positioning data for tourism statistics

Almost at the same time when Eesti Pank started to use mobile positioning data for assessing cross-border travel, Statistics Estonia started seeking for possibilities of extending the methodology for



making official tourism statistics according to Regulation 692/2011/EU concerning European tourism statistics. All other possible parties of cooperation in Estonia – Eesti Pank, OÜ Positium LBS and the University of Tartu – were also willing to cooperate. Despite a relatively extensive awareness campaign, the development action was not included in the statistical programme, which forms the basis for Statistics Estonia's work. So, no decisive actions other than the periodical exchange of information between the potential partners could be taken before 2012 when Eurostat launched a special grant project – Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics (Invitation to tender..., 2012). The project was carried out from January 2013 to March 2014 by a consortium consisting of OÜ Positium LBS and the University of Tartu, Statistics Estonia, Statistics Finland, the Institute for Tourism Research in Northern Europe (NIT), and the French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR). Eesti Pank could not participate in this consortium because of the heavy workload in its core activities.

The study was divided into five sub-studies, each of which concentrated on a specific aspect of using mobile positioning data for generating statistics on tourism flows: inventory of the use of mobile positioning data, access to the data, the methodology of data processing and the quality of the results, consistency with statistics from other fields, potential uses and benefits.

During the inventory, a total of 31 cases of using mobile positioning data in various fields in Europe and elsewhere in the world were found and studied. The results revealed that there is great interest in implementing new data sources (incl. mobile positioning data) for the production of surveys and statistics. Studying the cases of use shows the possible practical uses of mobile positioning data (Ilves, M., Karus, E., 2014).

The main findings of the study were as follows (Feasibility study..., 2014):

- Access to mobile positioning data is currently very limited, mainly because of the regulatory limitations. There are big differences between the EU countries.
- There is a need for a central framework for national statistical institutes (NSIs) and other stakeholders in order to obtain the data legally and according to an approved methodology so as to be able to produce comparable and reliable tourism statistics.
- Longitudinal data is a must for reliable tourism statistics in order to assess the whereabouts of the subscribers over a longer period of time (e.g. usual environment, differentiation of trips by length, identification of overnight stays, etc.).
- At present mobile positioning data can be used as a supplement rather than as a replacement source of data for the current official tourism indicators required in Regulation 692/2011/EU on European tourism statistics.
- Compared to the existing statistical processes, the use of mobile data as a source of tourism indicators provides improvements in several aspects, such as: timeliness (in some cases up to near-real time), access to statistical information previously not available (new indicators), calibration opportunities for the existing data, better resolution, and greater accuracy in time and space.
- Mobile positioning data can complement the currently used methods through mixed-mode data collection, enabling the sample size of the conventional survey to be decreased.
- Other domains of statistics, as well as disciplines outside official statistics, can benefit from joint processes and indicators generated via mobile positioning data, making the processing and use of mobile positioning data more cost-effective.

The main strengths of using mobile positioning data for producing tourism statistics were considered to be as follows (Feasibility study..., 2014):

- Fairly good consistency over time for the number of trips and nights spent compared to data based on 'traditional' methodologies.
- Superior coverage for overnight trips when compared to accommodation statistics because mobile positioning data can also cover trips in non-paid or non-registered accommodation.
- The option to compile more detailed breakdowns by region or country of residence.
- The option to apply common rules and criteria for identifying or classifying specific phenomena like the usual environment, secondary destinations, repeat visits, the frequency of visits, etc.



- The improved timeliness of statistics (up to near-real time) and the possibility of using mobile data for unconfirmed flash estimates.
- A higher level of automation in the production of statistics.
- The possibility of improving international cross-border statistics (mirror statistics) that reflect the travel network between different countries, provided that those countries use mobile positioning data for inbound and outbound tourism statistics.

The main weaknesses of using mobile positioning data for producing tourism statistics were considered as follows (Feasibility study..., 2014):

- A high level of complexity in accessing the data of mobile network operators and uncertain continuity in terms of accessing the data in the future.
- A relative lack of information on the purpose of the trip, expenditure, type of accommodation and means of transport used.
- Bias between some classifications due to the nature of the data, e.g. over-coverage of same-day trips due to the misclassification of overnight trips.
- Issues related to the qualitative understanding of tourism which can be misclassified due to not understanding the purpose of the trip (e.g. a respondent visiting relatives might not consider it as tourism, but it is classified as such according to quantitative standard international criteria).
- Over- and under-coverage issues related to the usage of mobile phones, e.g. tourists who do not appear in mobile positioning data at all, those who use several mobile devices or the roaming services of several MNOs.
- Difficulties with assessing the quality (especially accuracy) of statistics based on mobile positioning data because mobile phone usage during travel is largely unknown.

All in all, the study concluded that the possibilities of using mobile positioning data for statistics required by European Union (EU) Regulation No 692/2011 concerning statistics on tourism are rather limited. The main reason is the heterogeneity of the rules and regulations concerning access to mobile positioning data. The options of using mobile positioning data that were described in the project can be implemented only in those countries where access to the data has been granted. The main benefit of mobile positioning data comes in the form of new indicators and quick estimates.

The study also suggested that mobile positioning data opens up the possibility of calculating new indicators which are not required by current EU regulations, for example, in the following fields:

- Detailed analysis of tourism behaviour, including the frequency of visit for big events or cultural monuments.
- Distinguishing between repeat and first-time visitors assuming that longitudinal mobile positioning data are available.
- If personal data of the subscribers are available, they can be combined with the mobile positioning data to conduct a background analysis for the visitors or to study a certain segment of visitors.

4. Estonian official statisticians moving towards using new data sources

The feasibility study divided barriers to data access that require overcoming into the following categories (Feasibility study... 2014):

- Privacy and regulatory issues – how data can be accessed according to various pieces of privacy protection legislation.
- Public opinion – how the use of mobile data is perceived by the general public.
- Financial and business-related barriers — how and why MNOs should provide access to the data.
- Technological issues – what is technologically needed in order to be able to compile tourism statistics that are based on mobile positioning data and how the current data processing system can be amended so that the processing of the mobile positioning data is also supported by NSIs.
- Methodological issues – the quality and applicability of the principles of statistical production in relation to mobile positioning data.



The current legal instruments enforced in the EU do not clearly stipulate the form of processing mobile positioning data either for official statistics or for general use. Neither does Estonian legislation. Nevertheless, the Official Statistics Act of Estonia (2010) foresees that Statistics Estonia shall primarily use data collected in administrative records and databases, data generated in the course of the activities of state and local government authorities and legal persons or collected by them, if such data allow the production of official statistics complying with the quality criteria of official statistics. The owners of such data are obliged to provide Statistics Estonia with the data. It is not even required to inform individuals of the use of their personal data in the production of official statistics. The micro-data of a natural person shall be kept together with the personal identification code of the person during the validation of the micro-data. After validation, the personal identification code shall be stored separately from the other data of the person, ensuring the possibility for later linking of data.

In practice, it means that Statistics Estonia has access to mobile positioning data as any other big data if they are needed for completing a statistical action included in the statistical programme. Developing a methodology also counts as a statistical action.

In the case of Estonia, the main obstacle to using mobile positioning data has been a lack of resources – human resources for the development or adaptation of the methodology, and resources for obtaining the technology.

Data quality did not prove to be a big problem as during the above-mentioned international feasibility study access was gained to pilot data from three countries, but only the Estonian data proved to be of sufficient quality for the analysis (Feasibility study..., 2014).

One of the reasons behind the supportive public opinion on the access to all kinds of identifiable data by Statistics Estonia has been the well-defined strict rules of confidentiality and data protection. According to the annual survey carried out by an independent research company, Statistics Estonia was considered as the most independent and reliable government authority in 2014 (Survey on Communication..., 2015).

In 2015, Statistics Estonia joined the Infotechnological Mobility Observatory (IMO) together with the University of Tartu and Tallinn University of Technology. The aim of the IMO is to build a data infrastructure for mobility studies that integrates emerging (mobile phones, sensors) and traditional (individual level censuses, registers) data sources. The infrastructure to be generated will allow studying all aspects of population mobility and addressing the social challenges related to it on a qualitatively new level, both in terms of research as well as addressing the social challenges in Estonia. The IMO was established to raise the quality of scientific research of the changing patterns and forms of mobility and the related challenges.

There is also an idea to use mobile positioning data for the next round of the population and housing census (PHC) in 2020. The next PHC is planned to be carried out completely register-based in Estonia. Seventeen administrative data registers will be directly used for compiling the PHC variables, with some additional registers being used for quality checks. Nevertheless, a serious problem has been identified with the quality of residence data in administrative sources. Mobile positioning data could be a good data source for estimating and improving the quality of residence data. Firstly, unregistered emigrants can be identified by using mobile positioning and, secondly, the actual place of residence can be determined in the case of multiple items of residence information in administrative sources. The main prerequisite for realising this idea is getting identified mobile positioning data from the MNOs in order to link the data with other data sources.

In the shadow of the stormy fight for resources to use of mobile positioning data, Statistics Estonia has gained access to identified energy meter data of households. The data are already in Statistics Estonia's position and are planned to be tested for the construction of a model for distributing the electricity consumption of households on heating and cooling the housing, heating water, cooking and lighting. For getting background information on the number of household members and other details, the variables from social surveys will be linked to the smart meter data. The aim is to produce statistics according to the EU Regulation No 1099/2008 on energy statistics, as regards the implementation of annual statistics on energy consumption in households.



In 2014, Statistics Estonia had the first experience of using traffic sensor data. After a four-year break, the Foreign Visitors Survey was carried out in 2014. There was a need for auxiliary information in order to estimate the total number of foreign visitors for weighting survey data. There were good data sources for the number of border crossings at the airport, port and Russian border, but collecting data on border crossings at the Latvian border proved to be the most problematic. To solve the issue, it was decided to use traffic sensor data. Most of the road traffic to Latvia goes through two main roads – Tallinn-Ikla-Riga and Tartu-Valga-Riga. Close to both border points, there are permanent traffic sensors managed by AS Teede Tehnokeskus. They provided Statistics Estonia with traffic sensor data for the 10 vehicle categories with an interval of 15 minutes. Based on the data, the total number of visitors was estimated. The Foreign Visitors Survey data was used to estimate the share of foreign visitors among all visitors. Finally, the total number of foreign visitors was estimated using the total number of visitors and the share of foreign visitors (Tõnurist, A., Sõstra, K., 2015). Traffic sensor data was the only information source which enabled estimating road traffic between Estonia and Latvia. Despite minor problems with excluding domestic transport in Valga and the undercount of buses, we plan to use traffic sensor data for the Foreign Visitors Survey in the future as well.

5. Conclusions

Estonia is one of the first few countries where official statistics are published based on big data. One of the two producers of official statistics in Estonia – Eesti Pank – uses mobile positioning data since 2010 and publishes official cross-border travel statistics based on these data since 2012. Contrary to most EU Member States, there are neither any privacy nor regulatory issues in Estonia and the public opinion is supportive in general. Even so, due to the lack of resources, Statistics Estonia was not able to extend the methodology of tourism statistics. At last, in 2013–2014, thanks to the Eurostat grant project, Statistics Estonia participated in the international feasibility study on the use of mobile positioning data for tourism statistics. The study concluded that using mobile positioning data for statistics required by the EU Regulation No 692/2011 concerning European statistics on tourism are rather limited, but the main benefit of mobile positioning data comes in the form of new indicators and quick estimates.

Based on this information, Statistics Estonia has joined the Infotechnological Mobility Observatory together with the University of Tartu and Tallinn University of Technology. There is also an idea of using mobile positioning data for improving the quality of residence data during the next round of the population and housing census in 2020. Statistics Estonia has already had its first experience of using traffic sensor data and gained access to identified smart meter data, which are currently tested for identifying different types of energy consumption in households.

References

1. Ahas, R., Aasa, A., Roose, A., Mark, Ü., Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3): 469–486.
2. International Travel Statistics. (2015). Eesti Pank. [www] http://statistika.eestipank.ee/?lng=en#treeMenu/MAKSEBIL_JA_INVPOS/1410 (17.07.2015).
3. Invitation to tender for the supply of statistical services 2012/S 84-137069: Feasibility study on the use of mobile positioning data for tourism statistics. (2012). Eurostat. [www] http://ftp.infoeuropa.eu/rocid/database/000048001-000049000/000048731_2.pdf (17.07.2015).
4. Eurostat. (2014). Feasibility study on the use of mobile positioning data for tourism statistics. Reports on Eurostat Contract No 30501.2012.001-2012.452. http://ec.europa.eu/eurostat/c/portal/layout?p_1_id=6133613&p_v_1_s_g_id=0 (17.07.2015).
5. Ilves, M., Karus, E. (2014). Feasibility study on the use of mobile positioning data for tourism statistics. *Eesti Statistika Kvartalikirj. 2/2014*. Quarterly Bulletin of Statistics Estonia. [www] http://www.stat.ee/publication-download-pdf?publication_id=36396 (17.07.2015).
6. Kroon, J. (2012). Mobile positioning as a possible data source for international travel service statistics. Contributed paper to UNECE Conference of European Statisticians Seminar on New Frontiers for Statistical Data Collection. [www] <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/mtg2/WP6.pdf> (17.07.2015).
7. Mobiilposicioneerimisel põhinev väliskülastajate turismistatistikute uuring (in Estonian). (2009). OÜ Positium LBS, Eesti Pank. Tartu.
8. Official Statistics Act. (2010). Riigi Teataja. [www] <https://www.riigiteataja.ee/en/eli/ee/506012015002/consolide/current> (17.07.2015)
9. Liiklusloenduse tulemused 2014. aastal (in Estonian). (2015). Teede Tehnokeskus AS. Tallinn [www] http://www.mnt.ee/public/Aruanne_LL_2014.pdf (17.07.2015)
10. Survey on Communication between Government Agencies and Media in 2014. (2015). Turu-uuringute AS. [www] <http://www.stat.ee/dokumendid/178532> (17.07.2015)
11. Tõnurist, A., Sõstra, K. (2015). Foreign Visitors in Estonia. *Eesti Statistika Kvartalikirj. 2/2015*. Quarterly Bulletin of Statistics Estonia. [www] http://www.stat.ee/publication-download-pdf?publication_id=39431 (17.07.2015)