Mobile telephones and mobile positioning data as source for statistics: Estonian experiences

Rein Ahas[1], Margus Tiru[2], Erki Saluveer[2], Christophe Demunter[3]
[1]Department of Geography, University of Tartu,
Vanemuise 46, Tartu, Estonia, rein.ahas@ut.ee
[2]Positium LBS, Õpetaja 9, Tartu, margus.tiru@positium.com
[3]European Commission, DG Eurostat, tourism statistics section,
christophe.demunter@ec.europa.eu

Abstract

This paper introduces Estonia's experiences in collecting statistical data with mobile telephone positioning. In the globalizing world and Europe with opening borders, the movement of mobile phones is one of the easiest sources to record border crossings and traffic flows. In Estonia, Positium LBS and the Department of Geography of University of Tartu provide different statistical datasets derived from mobile networks with the specially developed software Positium Data Mediator. Today, mobile positioning is technically possible in most mobile networks; this is a very rapidly developing method which can be successfully used in different applications.

For generating tourism statistics, we use passive mobile positioning, a database of the locations of call activities performed by roaming (foreign) phones. We record the location and timing of call activities in network cells for every roaming phone. Each phone is described by its pseudonymous ID and country of origin. Data can be collected for incoming and outgoing tourism. For traffic studies and commuting studies, we use databases of domestic call activities and a special model which generates personal anchor points and OD matrixes. In the presentation, we describe the peculiarities of data, data gathering, sampling, handling of spatial databases and some analysis methods, using examples from Estonia.

The most positive aspect of mobile positioning is that nowadays phones are widespread and mobility data collection is more cost-effective than with previous data collection methods. Weaknesses of such databases are related to privacy issues, business secrets of mobile operators and peculiarities of data processing.

Keywords: Mobile positioning, tourism statistics, travel item of balance of payments

Prologue – tourism, travel and mobility statistics in the 21st century?

Since Adolphe Quetelet borrowed probability theory and statistics from astronomy to introduce these methods to social science around 1835, data collection has traditionally been based on censuses or sample surveys with citizens or enterprises as respondents.

In the past decades, several reasons have led to a rethinking of how the primary data for tourism and travel statistics can be collected in a more time-efficient and cost-efficient way, with the precondition of not jeopardising the quality of the statistics that make use of this primary data. Firstly, the abolishment of border controls within certain regions of the world has made it more difficult to capture visitors when they cross the border. The most straightforward example in Europe is obviously the free of movement of persons in the Schengen area.
Secondly, political or managerial decisions have recently called for a change in the production method for official statistics in order to make increasing user needs meet with budget constraints and growing concern about respondent burden. In this respect, the Commission Communication on the production method of EU statistics: vision for the next decade set out the principles for the way forward for official statistics in the European Statistical System (European Commission 2009).
Thirdly, promising research results presented at events such as Eurostat's *New Techniques and Technologies for Statistics* conference or the *International Forum on Tourism Statistics* has changed the statisticians' perception about the practical use of new technologies in their daily work.
Fourthly, statistical authorities have tried to find synergies between different fields of statistics. As a consequence, new technologies have a large potential of being suitable for many related fields of statistics simultaneously. Techniques developed for mobility statistics can also serve tourism or travel statistics, and vice versa. In the coming years, a combination of interacting sources will possibly create a little revolution for statisticians in these areas.

An important part of tourism and travel statistics and mobility surveys is the simple recording of *physical* flows.
The wider spread and reduced production cost of GPS devices will probably lead to the use of such devices to monitor mobility, travel, tourism during a short period (e.g. one or two weeks). Such method could be an alternative to the 'bookkeeping system' or 'diary' currently used. Possibly, this can be method or source to collect data on short trips of one or two nights or the – difficult – segment of same-day visits. First experiments in France gave very encouraging results (Armoogum J., Roux S., Marchal P. 2009).
To capture mobility, travel or tourism during a longer period, mobile positioning can offer an alternative to the traditional ex-post questionnaires in which respondents report on trips made during a specified reference period. The mobile positioning has been tested in Estonia, the experiences will be discussed in the next chapters of this paper.
GPS devices and mobile positioning can provide quantitative information in a cost-efficient and time-efficient way, but follow-up sample surveys may still be needed to collect additional, qualitative information on the trips (e.g. means of transport or accommodation). However, these labour intensive sample surveys could be based on much smaller samples than is currently the case.
Another important part of tourism and travel statistics, relates to the *monetary* flows. In recent years, statistical offices have been exploring the possibilities to use credit card information, be it as primary data source or as auxiliary information for calibrations or quality control of sample-based information.

The fact that research experiments in certain countries showed encouraging results, proves that the use of new technologies is not just another flavour of the month or wishful thinking. We are rather confronted with a self-fulfilling prophecy: the belief in the new technologies can actually influence the behaviour of the statisticians. At least, it should be food for thought for official statisticians, and not only in the domain of travel, tourism or mobility.
In this intellectual exercise the opportunities as well as the threats need to be considered.
The application of new technologies can make statisticians, justly, dream about improved timeliness (through a reduction in collection and processing time), improved data quality (through a lower recall bias, a lower rate of data entry errors or an increased consistency and harmonisation) and a reduction of burden on respondents and administrations.
There is of course another side to the medal. A number of potential risks can turn the dream into a nightmare. Obvious issues are the privacy rules and the cost for accessing to the databases, the continuity of the access and the possible systematic bias (e.g. use of foreign sim-cards, mobile phone penetration rates). Further, a number of methodological issues need to be solved. Firstly, existing data can include a huge amount of information, but not necessarily with sufficient detail. Secondly, finding the relevant travel or tourism flows in an enormous database of roaming data or GPS registrations is not straightforward (not to mention the key tourism statistics criterion of the delimiting the "usual environment" for which longitudinal data may be required!). Besides the challenges in terms of access or methodology, a last factor can play an important role, namely trust. Indeed, when applying new technologies to obtain better, faster and cheaper statistics, not only the users will need to be convinced of the robustness of the data but also the official statisticians themselves will need to make a mental switch and be ready to depart from the data collection methods that have been used for the past 200 years.

1. Introduction

In this paper, we introduce mobile (telephone) positioning based methods for collecting statistical data about human movements. Mobile telephone positioning is often considered to be a novel and exciting source of information for studying the spatial dynamics of human society. Today, mobile positioning is technically possible in most mobile networks; this is a very rapidly developing technology which can be successfully used in different applications.
In Estonia, Positium LBS and the Department of Geography of University of Tartu provide different statistical datasets derived from mobile networks for government organisations with the special software Positium Data Mediator:

a) inbound and outbound tourism statistics (Ahas et al 2008; Tiru et al 2010)
b) travel item of the balance of payments (Positium 2009)
c) transportation flows and OD matrixes (Saluveer & Järv 2009)
d) everyday mobility and commuting (Silm & Ahas 2010).
e) personal and common anchor points (Ahas et al 2010)

In the globalising world and Europe with opening borders, the movement of mobile phones is one of the easiest sources to record border crossings and traffic flows. Major weaknesses of such databases are related to privacy issues, business secrets of mobile operators and data processing peculiarities. The database is large, requires special GIS

referencing, temporal and spatial interpolation and is supported by complicated data handling software.

2. Methodology

Mobile positioning is tracking the location of mobile telephones. Generally, it can be divided into active and passive mobile positioning. Active mobile positioning is used for tracking the location of mobile phones in real time using mobile positioning system (MPS) (Ahas et al. 2008). There are many technical solutions for active real time tracking of telephones. The cell identity method determines the network cell where the telephone is located. More complex and precise methods use trilateration by obtaining the distance from many antennae using direction and time lag of signals. Most of the smartphones with built-in GPS devices and innovative networks use Network Assisted GPS (A-GPS) positioning and navigation services which allow positioning with GPS quality and independently of operators will and charges.
Location data from passive mobile positioning is automatically stored in memory or log files of mobile operators (Ahas et al. 2008). Operators' systems generate very large amount of mobile phones usage data with a lot of location information attached to it. Mostly this data is used internally for business and marketing purposes by network carriers: charging clients for services, providing usage statistics, marketing purposes etc. Location data can be described as by-product and used seldom. This data also holds valuable information for generating anonymous statistics about space-time movement of phones (phone users) cost-effectively.

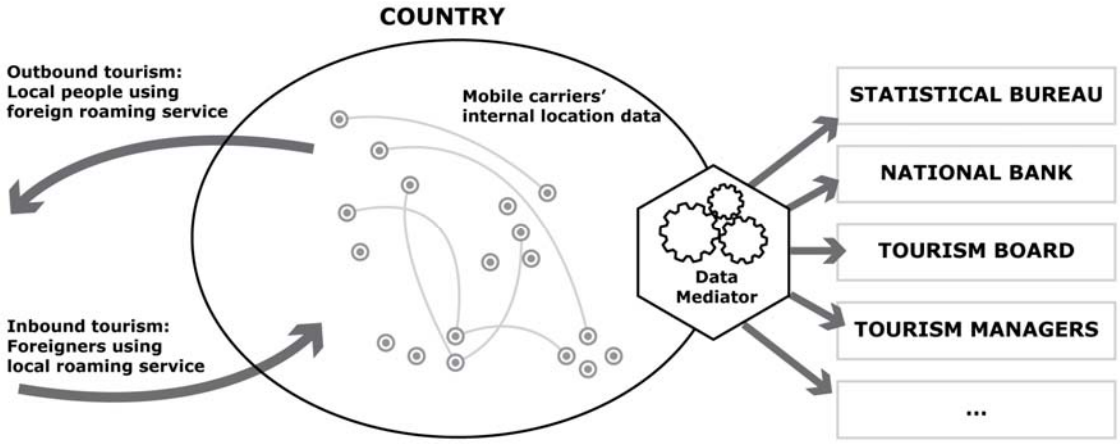

Figure 1. Inbound and outbound tourism data flows in Positium Data Mediator.

One use of such data is generating data for tourism statistics, tourism studies and balance of payments statistics as introduced in this paper. Passive mobile positioning data, more precisely roaming service usage inside and outside the country, represents a new approach for gathering inbound and outbound tourism statistics (Figure 1).
The specially developed program Positium Data Mediator intermediates anonymous and pre-processed statistical data from strictly protected systems of mobile network operators

and provides output data feed for statistical bureaus, national banks and other users who use such data in addition to traditional data. Because of privacy and data protection issues, the generated statistics have limited features for researchers.

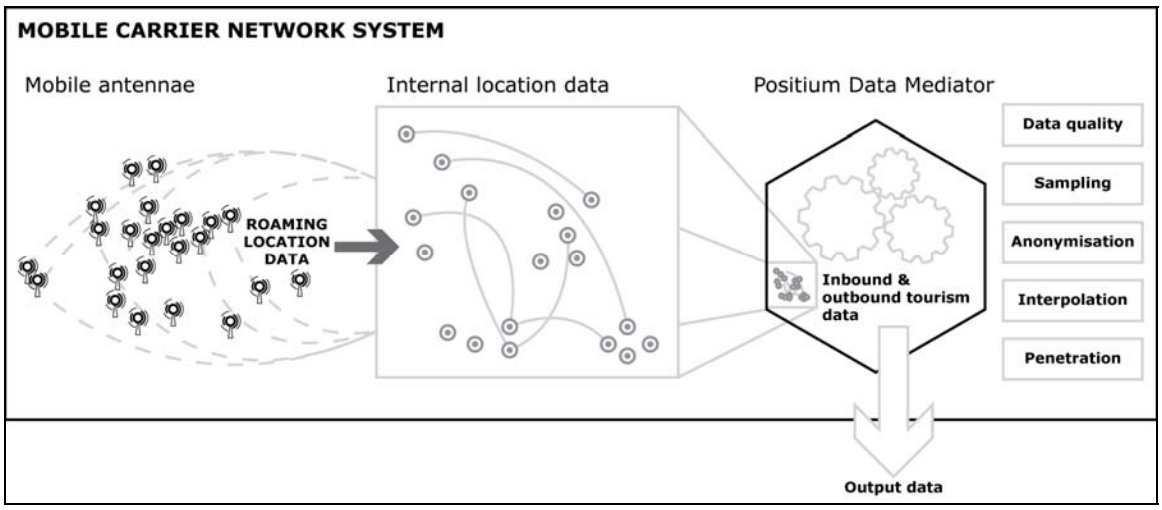

Figure 2. Structure of Positium Data Mediator inside mobile carriers' systems.

In terms of the utilization of mobile positioning data, several significant obstacles can be brought up, the most complicated one of which is surely the collaboration with mobile operators who, in essence, are interested in selling their reliability and confidentiality rather than their data. There are also many technical, methodological and legal issues that need to be overcome to collect mobile positioning data. The collaboration between Positium LBS and the Department of Geography of University of Tartu has resulted in the creation of the special data mediation software called Positium Data Mediator that can find solutions to a big part of the previously mentioned issues. Positium Data Mediator is a software that is partially located in the mobile operator's system, under the operator's control but it is partly also controlled by a data mediator. A data mediator can be an independent body, an operator as well as an end-user, who, for a certain reason, needs to acquire such data from the operator. In Estonia, for example, Positium LBS acts as a data mediator and the company has therefore also developed the software to perform the task.

In the operator's system and under the operator's control, the Positium Data Mediator software is used to select data and protect business secrets and personal. During these activities, the respondents are given randomly selected pseudonymous IDs and the future identification of mobile phone users will not be possible. Along with this, a unified data stream out of the operator's system is being prepared. The operator does not need to fear damage to people's privacy or business secrets caused by the collected data.

Firstly, the data mediator will carry out a quality control on the data gathered from the operator's system. As the amount of data is huge, filters to find and correct errors were developed, based on the characteristics of the data. The filters were elaborated during various process-oriented research activities. It is not possible to control and correct such capacious databases manually. Secondly, the application of a data sampling model is to be performed. In order to ensure representative data and scientific quality, the positioning data has to be specially evaluated and processed. In order to do so, the data from various

studies, validated databases and three specially developed models are used. The next step is to spatially interpolate the data. The raw mobile positioning data is not connected within any administrative to borders in any way. To make it possible to statistically describe administrative units of various hierarchical levels, the data needs to be spatially interpolated. The spatial interpolation will also have to ensure chronological comparability, so as to evaluate trends and changes in terms of people. The Positium Data Mediator uses a special GIS module to perform that step. The last part of the data collection system is its adaption to the needs of various data consumers. Based on their characteristic needs, the system has a separate query window for each consumer group. In some fields, consumer-friendly query environments have been developed (e.g. Positium Tourism Barometer and also a special interface for the Bank of Estonia).
When processed as described above, the researchers can use standardized data, the quality and representability of which has already been controlled. The system has been programmed to use various data inputs and in order to calibrate the data, special on-demand surveys are carried out twice a year. The whole system is continuously evolving with the help of a variety of surveys and use cases. The existence of an integrated system that takes the needs of different operators into account may be an important factor in engaging operators in the surveys. In the present time, the data in the system is renewed at the end of each week or month. The Estonian Ministry of Internal Affairs, for instance, ordered an evaluation of the emergency situation in relation to the snowstorms in December 2010. The data was used to specify the number of people in need of help and to contact those people. It is foreseen that in about 1-2 years time Positium Data Mediator will also start reflecting real-time data, which means that governmental authorities can evaluate certain processes and indicators on the basis of real-time maps. This will change the way statistics are collected and analysed so far.

3. Methodological problems and sampling issues

Using mobile phones for studying spatio-temporal behaviour of people and gathering statistics using this technique has several methodological peculiarities. Wide distribution of mobile phones and their active use is definitely a positive aspect of this methodology. Mobile phones are widespread in developed and developing countries; the usage is relatively widespread among different income, age and other social groups of the society as well as geographically (depending of course on network coverage and density). This enables to collect data more easily and extensively. Cost-effectiveness is certainly a positive aspect of this method, since the operators have recorded the data automatically, there are no direct costs for recruiting respondents and getting in touch with them, which forms a material part of the budget in ordinary surveys. Automatic data collection also serves as a substantial advantage, because compared to inquiries and travel diaries this method does not entail the problems of human forgetfulness and recollection of data. Compared to research based on GPS, the advantage of mobile positioning is connected with battery drainage: in GPS based surveys the quick discharging of battery serves as the main problem and people are often not motivated to recharge the battery. However, the mobile phone battery is recharged meticulously, since the phone is one of the essentials and it is always carried along.

Important aspect of mobile positioning data is its speed. As data is collected automatically, it can also be processed automatically and statistical results delivered close to real-time. This creates possibilities for fast analysis and quick response time. Traditional statistics and studies take a lot of time and often produce results slowly. Therefore the real-time approach is without any doubt the positive aspect of this methodology.

The most substantial problems of mobile positioning are those related to privacy. This topic will be handled below, but Positium Data Mediator and different authors (Reades 2010) offer several solutions to solve this problem.

Another important source of problems is sampling. Although mobile phones are widespread, there are not enough researches conducted about their utilization and too few generalizations have been made. Who, where and when are using mobile phones and for what reasons? It is essential to know the usage pattern of the mobile phones in systems which receive the telephone usage data as input. For instance, the essence of data used for generating tourism statistics. In the Estonian case location information is retrieved from various call activities of users. Therefore it is essential to understand the logic of phone usage by different user groups as it is apparent that different age and social groups, different income groups, nationalities and tourist types use mobile phones differently. But no accurate related scientific studies have been conducted.

Research conducted in Estonia shows that the number of call activities (i.e. number of registered locations) varies from 2.1 (Japan) to 10.2 (Azerbaijan) averaging 4 location points per day for all tourists. People with a higher income make more phone calls, for example, the inhabitants of Tallinn, the capital of Estonia, make on average 6 to 8 calls per day, the inhabitants of Tallinn suburbs, who are financially even more successful, make approximately 8 to10 outgoing calls per day. People who live in rural areas and have smaller income make fewer phone calls. The analysis also shows that according to age groups, the most active callers are people between 20 and 25 years, the number of calls decreases with age.

The method of study and results are also affected by the location where the phone is used. For example, the error of determining the anchor points of home and work is the biggest, when the phone is being used when moving around and in means of transport, which is a habit for many people (Ahas et al 2010). Tourist events and attractions also have a negative impact on data collection, because people do not use their phones in these events and often switch off the phones. Our experience with tourism statistics indicates that generally tourists do use mobile phones, figure 4 shows the visiting days of 10 visitor countries and the average number of phone calls in Estonia in 2010.

Tourism statistics generated from mobile positioning data differs in many ways from official statistics. The reasons are majorly in methodologies and the essence of the data. As seen in figure 3 inbound statistics on country level can differ a lot from official accommodation data as mobile positioning data includes also many tourist groups as one-day visitors, transit visitors and visitors who don’t stay in official hotels that provide official statistics. The correlation between official statistics and mobile data in Estonia is higher for nations that act like Germany with majority of visitors staying at official hotels for many nights and lower for close neighbours who make occasional one day visits for shopping or leisure and are not registered in official statistics.

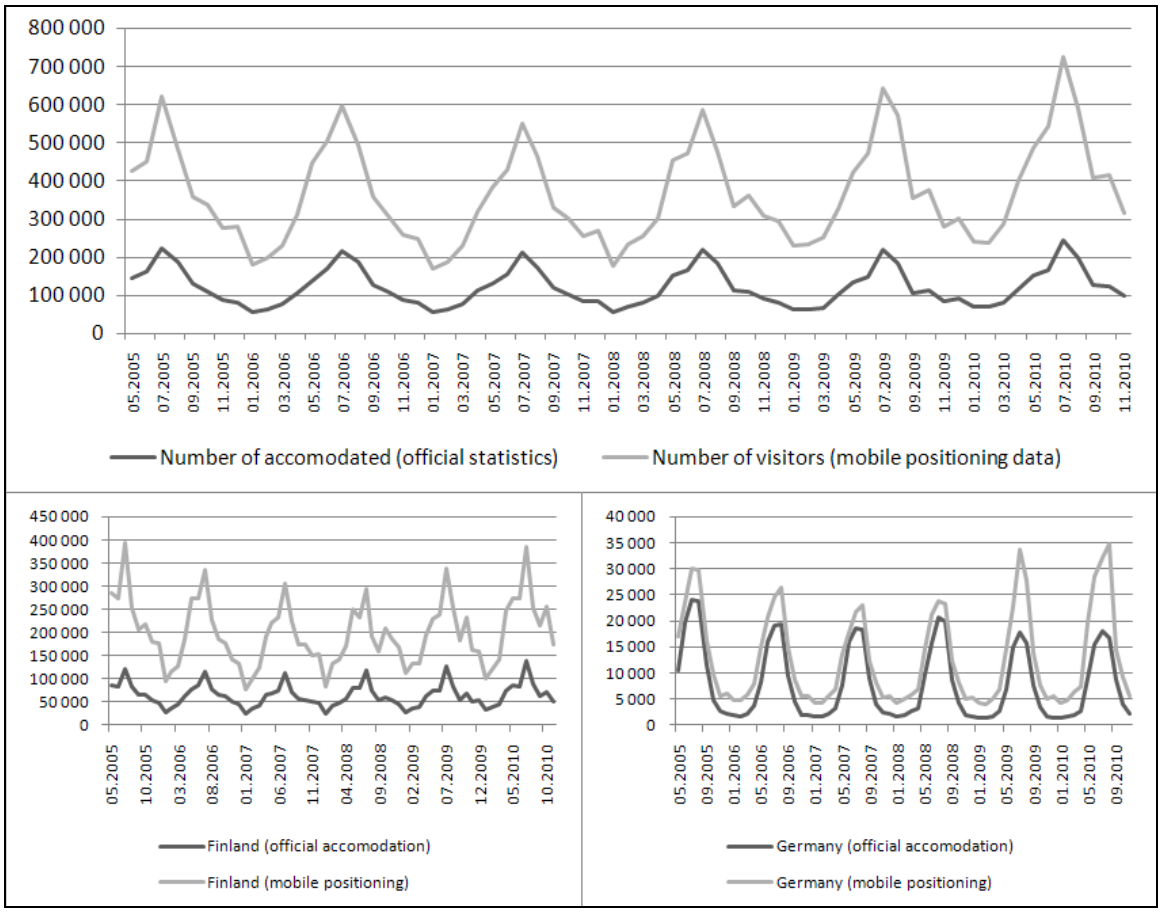

Figure 3. Comparison of official and passive mobile positioning based inbound tourism statistics for all tourists, Finns and Germans.

The following important issues that influence the usage of mobile phones are related with the costs of data handling. The data volume of positioning data is so large, that the handling and utilization are quite complicated. Errors are also difficult to detect and correct in large databases. Due to that the additional computing costs and labour costs related to system administration have to be taken into account.

Despite several methodological and technical weaknesses the data of passive mobile positioning is geographically and temporally more accurate than most of the conventional tourism statistics. For example, the accommodation data is usually available with monthly accuracy and locations are only determined with county or municipality level accuracy, moreover the data only covers the segment of tourists staying at accommodation establishments while less than half of the tourism trips made by European residents is spent in this type of accommodation (European Commission 2010). But mobile positioning provides data with one second precision and antenna location accuracy, it enables to conduct more adequate research and use data in very different applications, for instance, in marketing and destination development (Kuusik et al 2010). Due to that, the demand and interest for statistics based on mobile positioning is increasing.

4. Utilization of mobile positioning data in Estonia

Mobile positioning data has been used for studying the time-space behaviour of people and tourism in Estonia since 2001, when the scientist of the Department of Geography of University of Tartu, the planners of Hendrikson & Ko and the architects of Urban Mark jointly established a spin-off company Positium LBS. Since then the positioning data has been used in various projects, research and art.

In the field of urban planning the connection between the Tallinn city centre and other city districts has been investigated as well as the influence of suburbs to the development of the centre (Ahas et al 2008). The results indicate that the city centre of Tallinn is tightly integrated in the activity spaces of the residents of new communities of Tallinn and the expansion of Tallinn may rather be considered as the expansion of a central city and not so much as suburbanisation (Tammaru et al 2009). The new city districts have not developed suburban space usage and lifestyle. The urban planning projects of Tallinn have also used the data to map the routes and stopping points of the tourist flow emanating from the port and monitor the change in city functions. The projects of Tallinn also participated in the Venice Architecture Biennale in 2006.

The data has been quite extensively used to assess the influence of peripheral and influential areas. Everyday routes and movement between anchor points are easily detectable by passive mobile positioning. It is possible to monitor the location of a substantial part of the population and their movement at different moments. The most exhaustive project is “Survey of Regional Pendulum Migration” (“Regionaalne pendelrändeuuring”) (Ahas et al 2010b), according to which the country’s regional planning is organised and which has caused the initiation of creation of a monitoring environment based on state information system. The monitoring environment should develop into an opportunity to estimate the location of the society and the changes in real time. This information is used to direct public events and to grant the internal security of Estonia.

Important application areas of positioning data are surveys of transport and travel behaviour. The developed model of anchor points and the related methods for generating origin-destination matrixes enables to gain information about the transport needs and traffic flow much more cost effectively (Saluveer & Järv 2009). The Scandinavian road-planning enterprise Ramboll has used this to compile the biggest road project in Estonia – the reconstruction project of Tallinn-Tartu road. Tartu Eastern roundabout has been engineered according to the similar study. The public transport systems of several regions have been developed based on this information and the impact of urbanisation to traffic flow has been studied (Ahas et al 2009). The vast value of the data entails a possibility to easily calculate the geographic distribution of traffic and to connect the everyday traffic flows with actual destination points.

Many surveys and application projects have been conducted in the field of tourism and the data has also been noticed and used in scientific literary works (Ahas et al 2007; Ahas et al 2008). Surveys have been conducted to investigate the impact of public events to local life and the number of visitors, seasonality of tourism and the formation of tourism routes in destination points. A monitoring and inquiry environment Positium Tourism Barometer has been developed for such purposes and it is being used by several Estonian local governments, government authorities and science groups in order to conduct their

surveys (Tiru et al 2010). The most active users of the environment are the local governments engaged in tourism development, who want to observe the trends of time-space behaviour of different visitor groups. Different consultation companies also make use of the data by preparing the analyses and project applications in order to develop tourism in Estonia. The tourism barometer is also being used by funds financing tourism projects, which helps them to evaluate the reality of the projects and post-monitor the effects of the financial investment. Another fast-growing data application field is the marketing of tourism and the destination location. Mobile positioning offers innumerable possibilities in this field and for marketing organisations this data has opened new perspectives for development of marketing conceptions (Kuusik et al 2010).
Additionally, tourism statistics is also used by consumers of general statistics, above all the Bank of Estonia, who uses the statements of Positium Tourism Barometer for already 3 years in order to calculate the country's balance of payments according to outgoing tourism data (travel item). Their active usage of the database can be accounted for the fact that they can evaluate the incoming and outgoing visitors in a situation where there is no border statistics and the census survey with adequate volume is quite expensive.

5. Summary

This paper introduced Estonia’s experience in generating movement statistics with the help of mobile positioning. New methods always induce methodological questions and discussions. The development of Estonia’s system, together with methodology and technology, has happened gradually since 2001. The development has been scientific and conducted in cooperation with the Department of Geography of University of Tartu and EMT, an Estonian mobile operator, which are the reasons why the consumers have started to trust the system. Today all the technical and methodological wisdom has been centred in special software called the Positium Data Mediator. This program collects the information of different operators, protects personal data and business secrets and standardises the outputs. The users of such standardized output are several Estonian organisations involved in tourism, planning, transport and regional development. The data is also being actively used by a fast-growing community of scientists, because various researchers need this kind of information. The future possibilities of mobile positioning involve the collection of real-time statistics and the development of monitoring systems based on the analysis of this data, which can change the overall trends in the field of statistics.
Despite the positive examples of Estonia, mobile positioning data involves several problems. For instance, there are many unanswered questions related to the sample and quality of the data. Protection of data and privacy also serves as an important issue that is continually in the centre of attention. The example of Estonia can serve as a positive basis to promote the gradual conduction of similar surveys in other countries. In globalising world and Europe with open borders the handling of new data collection methods is essential and the experience of different countries helps to develop and harmonise the systems.

References

Ahas, R., Mark, Ü. 2005. Location based services – new challenges for planning and public administration? Futures, 37(6): 547-561.

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.

Ahas, R., Silm, S., Järv, O., Saluveer E., Tiru, M. 2010a. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones , Journal of Urban Technology, 17(1): 3-27.

Ahas, R., Silm, S., Aasa, A., Leetmaa, K., Tammaru, T., Saluveer, E., Järv, O., Aasa, A., Tiru, M. 2010b. Regionaalne pendelrändeuuring. Tellija: Siseministeerium. Study report: Regional Commuting Study, Estonian Ministry of Interior. 264 p.

Armoogum J., Roux S., Marchal P. 2009, *"Acceptability of the use of new technologies by interviewees in surveys",* presented at the NTTS 2009

European Commission 2009, Communication COM(2009)404 on the production method of EU statistics: vision for the next decade.

European Commission 2010, *"Tourism statistics in the European Statistical System: 2008 data"*, Eurostat Methodologies and Working Papers.

Kuusik, A., Tiru, M., Varblane, U., Ahas, R. 2010. Process innovation in destination marketing: use of passive mobile positioning (PMP) for segmentation of repeat visitors in case of Estonia, Baltic Journal of Management (submitted).

Positium LBS 2009. Mobiilpositsioneerimisel põhinev turismistatistikute uuring. (The study of mobile positioning based tourism statistics). Study Report, Tartu, 37 p.

Reades, J. 2010. Finite State Machines: Preserving Privacy When Data-Mining Cellular Phone Networks, Journal of Urban Technology, 17(1): 29 – 40

Saluveer E. and Järv.O. 2009. Measuring traffic flows with passive mobile positioning data. Working paper of Positium LBS 2009.

Silm,S., Ahas, R., 2010. 'The seasonal variability of population in Estonian municipalities, Environment and Planning A, 42(10) 2527-2546.

Tammaru, T., Leetmaa K., Silm, S, Ahas, R. 2008 New residential areas in the Tallinn Metropolitan Area. European Planning Studies, European Planning. Stud 17(3): 423-439

Tiru, M., Saluveer E., Ahas, R., Aasa, A. 2010. Web-based monitoring tool for assessing space-time mobility of tourists using mobile positioning data: Positium Barometer. Journal of Urban Technology, 17(1): 71-89.