Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

Consolidated Report

Eurostat Contract No 30501.2012.001-2012.452

30 June 2014

eurostat

NIT NEW INSIGHTS FOR TOURISM

UNIVERSITY OF TARTU 1632

ES statistics estonia

positium

IFSTTAR

Statistics Finland

Eurostat

Europe Direct is a service to help you find answers to your questions about the European Union.
Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

Consolidated Report

Eurostat Contract No 30501.2012.001-2012.452

30 June 2014

Authors of this report (in alphabetical order):
Rein Ahas (University of Tartu), Jimmy Armoogum (IFSTTAR), Siim Esko (Positium), Maiki Ilves (Statistics Estonia), Epp Karus (Statistics Estonia), Jean-Loup Madre (IFSTTAR), Ossi Nurmi (Statistics Finland), Françoise Potier (IFSTTAR), Dirk Schmücker (NIT), Ulf Sonntag (NIT), Margus Tiru (Positium)

More information on the European Union is available on the Internet (http://europa.eu).
Luxembourg: Publications Office of the European Union, 2014
ISBN 978-92-79-39762-2
doi:10.2785/55051
Printed in Luxembourg

Foreword

Statistics are not static, data series are — obviously — a dynamic phenomenon. However, the back-office of statistics is also constantly shifting. On the one hand, user needs for data increase and, on the other hand, the production methods to collect and compile statistics are not in a vacuum but exposed to exogenous events or evolutions.

Although the legal framework for European statistics on tourism (Regulation 692/2011) has fostered the setup of a system of harmonised and comparable tourism statistics that succeeds in addressing a wide range of user needs, some shortcomings do persist. Users need new data or more complete data and users need data much faster than currently released.

Three recent changes are impacting the production of tourism statistics: a changing geopolitical environment (putting in question, for instance, the continuity of border surveys), a changing technological environment (new tools and devices have entered every citizen's daily life) and a changing working environment for statisticians (how to meet expanding user needs with shrinking budget resources?).

To seize the momentum of newly arising data sources as a facilitating factor to answer these changes, Eurostat launched in 2012 a Call for Tender for a **feasibility study on the use of mobile positioning data for tourism statistics**. The drivers for this project were analogue to those embedded in the higher-level ESS (European Statistical System) Vision 2020: data revolution, new metrics and the cost of statistics.

There were five main objectives of the project: to assess the feasibility to access databases with mobile positioning data in European countries, to assess the feasibility to use such data for tourism statistics, to identify and address the main methodological barriers to implement this data source as a building block for official statistics, to assess the potential impact on cost-efficiency of data production and to assess the possibility of expanding the methodology to other domains and define joint algorithms. The set-up of the project was a mix between theoretical and practical or empirical work.

The motivation to launch such an initiative at the supranational European level stemmed from the aim of, in the end, producing harmonised figures and the possible economies of scale as compared to regional or national initiatives. Indeed, the challenges and barriers to overcome are often independent of the country (at least within the European Union) and independent of the area of statistics. The different reports resulting from this project have the ambition to address most technical, legal or quality-related questions that may surround the use of mobile positioning data for official statistics and thus to avoid a duplication of work within the ESS.

We hope that the outcomes of this feasibility study — which was also one of the first concrete actions in the context of Eurostat's explorations of big data — will be inspirational for statisticians, researchers and users that are fascinated by the potential of this relatively new data source.

The current report is the consolidated report of all tasks executed by the international consortium that conducted the feasibility study. This report tries to summarise in a nutshell the tasks and findings of the study and to guide the reader to the five more detailed reports that cover each component of this feasibility study.

Christophe Demunter

Unit G-3 'Short-term business statistics and tourism', EUROSTAT

Contents

Summary

Overview

The aim of the current study was to assess the feasibility of using mobile positioning data for generating statistics on domestic, outbound and inbound tourism flows, and to address the strengths and weaknesses related to access, trust, cost, and the technological and methodological challenges inherent in the use of such a new data source. The international consortium that conducted the study concentrated on the various aspects involved in the use of mobile positioning data in terms of tourism statistics and other domains: an overview of the situation involving the use of mobile data; accessibility to the data from the legal, technological, financial and business aspects, including possible cost and burden implications; methodological principles of statistical data collection and compilation, including evaluation by using different quality aspects and comparing the results against existing traditional methods; opportunities offered by, as well as limitations inherent in, the use of the data source. A more detailed discussion on each of these aspects can be found in the detailed reports, which can be consulted and downloaded from the **Eurostat website**.

The study demonstrated extensive interest in this data source and possibilities for a wide range of uses when it comes to tourism statistics and other domains, while at the same time acknowledging the multiple problems associated with accessing and processing the data. The number of projects and research areas that are focusing on the use of mobile positioning data is constantly increasing. Tourism statistics is one of the domains in which the opportunities are rather clear as the properties of the data correspond to the nature of the tourism activities. Inbound, outbound roaming and domestic data stored by mobile network operators (MNO) clearly corresponds to the respective inbound, outbound and domestic domains of tourism, however, not without some methodological reservations. The use of mobile positioning data (and also 'big data' in general) by users of tourism statistics includes the following expectations:

- Reducing both the burden on the reporting units and the cost involved in statistical processes by (fully or partially) replacing the existing, relatively expensive methods with new data sources;
- Expanding the available options in terms of measuring tourism activities through new indicators;
- Improving timeliness;
- Improving time and spatial accuracy.

The overall opinion and expectations when it comes to the use of mobile data are rather high, and it is generally believed that this data source will increasingly be implemented in the following years in terms of tourism activities and in other domains. National Statistical Institutes (NSI) perceive this data source mostly as complementary, and in some cases it is also seen as a potential replacement for existing data sources and methodologies.

Main Findings of the Study

- Access to mobile positioning data is currently very limited mainly because of the regulatory limitations. There are big differences between the EU countries.
- The study concludes that there is a need for a central framework for NSIs and other stakeholders in order to obtain the data legally and according to an approved methodology in order 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 the trips by length, identification of overnight stays, etc.).
- Based on the outcomes of this study, it can be concluded that 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.
- However, the use of mobile data as a source for tourism indicators introduces several aspects of improvement compared to the existing statistical processes, such as: timeliness (in some cases up to near-real time), access to statistical information previously not available (new indicators), calibration opportunities for existing data, better resolution, and 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 spheres 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 mobile positioning data:	The main weaknesses of mobile positioning data:
1. Fairly good consistency over time for the number of trips and nights spent compared to data based on 'traditional' methodologies; 2. 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; 3. The option to compile more detailed breakdowns by region or country of residence; 4. The option to apply common rules and criteria for identifying or classifying specific phenomena like usual environment, definition of secondary destinations, repeat visits, frequency of visits, etc. 5. Improved timeliness of statistics (up to near-real time) and possibility to use mobile data as unconfirmed quick indicators. 6. Automation level of statistical production; 7. Possibility to improve cross-border international statistics (mirror statistics) that reflects the travel network between different countries provided that those countries use mobile positioning data for inbound and outbound tourism statistics.	1. Complexity of the access to MNOs' data and uncertain continuity of access to the data in the future; 2. Relative lack of information on the purpose of the trip, expenditure, type of accommodation and means of transport used; 3. Bias between some classifications due to the nature of the data, e.g. over-coverage of the same-day trips due to misclassification of overnight trips; 4. Issues related to the qualitative understanding of tourism that can be misclassified due to no understanding of the purpose of the trip (e.g. visiting relatives might not be considered as tourism if the person was asked, but it is classified as tourism according to quantitative standard international criteria); 5. Over- and under-coverage issues related to the usage of the mobile phones, e.g. the tourists who do not appear in mobile positioning data at all, who use several mobile devices or who use the roaming service of several MNOs; 6. Difficulty to assess the quality (especially accuracy) of statistics based on the mobile positioning data because mobile phone usage during travel is largely unknown.

Feasibility of Access, Use and the Opportunities in Mobile Positioning Data

There are many **issues with data access** that require resolving. Generally, the barriers can be divided into the following categories:

- 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 should MNOs 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 can the current data processing system 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 and the Member States do not clearly stipulate either the specific conditions or the form for processing mobile positioning data for official statistics and for general use.

The main principles covering the subject stipulate that if mobile data can, initially, be classified as being directly or indirectly identifiable (e.g. personal data), then any processing of such data may be conducted either

(a) bearing in mind the fact that the subscribers have given permission for such processing;

(b) national legislation allows or compels on the one hand the MNOs to provide the data and on the other hand, the NSIs to process it for official statistics; or

(c) for the internal purposes of MNOs that are required for their commercial activities.

If the data and the processing can guarantee anonymity (meaning that none of the subscribers can be directly or indirectly identified), then processing is allowed for any purpose.

Although these conditions are interpreted differently depending on the country, it can be concluded that the best option for NSIs would be to compel the MNOs through legislation to provide data either by directly transmitting raw data or by processing and transmitting the resulting statistical indicators. However, this could also be a circular argument. In order to implement the statutory power to obtain the data, NSIs need to conduct a pilot to determine feasibility. At the same time, the MNOs have no legal footing to provide the data in the first place.

The easier option would be then to apply an anonymisation algorithm to the raw data before processing. However, currently, there are no conclusive methods for anonymising data while preserving the necessary longevity of subscriber-based data (longitudinal data).

Besides the regulatory issues, **public opinion** is a factor to be taken into consideration. This data source is highly sensitive and although the main idea behind using this data for collective statistical purposes, and thus not focusing on the individual profiling of the subscriber, is by itself understandable and noble, it can create perceptions of people being tracked. Therefore, the use of the data for purely statistical and research purposes should be conducted in parallel with public-awareness-raising activities concerning the purpose and methods of processing of the data.

Mobile Network Operators (MNOs) realise the potential of using this data source but tend to take a conservative stand as they see many obstacles, mainly in regulatory and privacy protection, as well as business and financial aspects that could jeopardise their core business. They are looking for practical solutions that would enable them to provide the data for statistical purposes and possibly generate new revenues by selling big data.

The **technology** for processing large amounts of data for statistical purposes is quickly developing and technically the implementation of the technology does not present any serious obstacles besides the cost for the hardware, software and maintenance, which depends largely on the desired timeliness and the volume of data. The cost of the system depends on whether the MNOs have to process the data into aggregated statistical indicators before transmitting to NSI or just extract and transmit raw data to NSI for processing. The overall cost of the first system is obviously higher when compared to the latter.

The processing steps for compiling tourism statistics based on mobile positioning data are different than those used in the traditional methods. The **methodology** might differ depending on the type of data provided by MNOs. The basic type of data that is the most easily accessible by MNOs is the dataset of Call Detail Records (CDR) that represents phone activity — calls, messaging. Alternative data types include Data Detail Records (internet usage), location updates (periodical identification of a device within the network antennae) and others. These data sources are more detailed and accurate in terms of frequency of the events and geographic accuracy, but they are often not available or require enhancement of the technology by MNOs.

This study concentrated on the possibilities of longitudinal data that presents the best methodological options for compiling tourism statistics. The use of non-longitudinal initial data, such as aggregated data or data with temporary subscriber ID [1] limits the possibilities of generating reliable statistical indicators drastically, mainly because it is impossible to measure the duration (and frequency) of the stays of subscribers in specific places, which is crucial for identification of a subscriber's country of residence and usual environment [2].

The data processing from the initial raw data up to resulting statistical indicators includes data extraction, frame formation, data compilation, estimation and combination of data from several MNOs. From the perspective of methodology, the most crucial step includes the identification

[1] **Subscriber identification codes that are changed over a period of time.**

[2] **A geographical area, though not necessarily a contiguous one, within which an individual conducts their regular life routines (Eurostat 2013).**

of the usual environment of the subscribers from different datasets (inbound roaming, outbound roaming and domestic data) to differentiate between regular trips from tourism trips where complex phenomena like cross-border commuting, long-term visits and pass-through transits have to be delimited. In the sphere of tourism statistics the resulting **data can provide the following indicators and breakdowns**:

- Indicators: number of trips/visits; number of nights spent; number of days spent; number of unique visitors.
- Breakdowns (or classifications): country of residence/place of residence; aggregation of time (day, week, month); aggregation of space (different level of administrative units, grid); duration of trip/stay (same-day/overnight trip); main destination, secondary destination, transit pass-through; collective movement patterns; repeat visits.

Based on the analysis of the methodology, it can be concluded that mobile positioning data, at present, **cannot be a replacement, but rather a supplementary source** for the official tourism indicators required in the current Regulation 692/2011 concerning European statistics on tourism. An important reason is that the Regulation is often based on specific data collection methods (e.g. accommodation statistics) and that mobile positioning data does not reflect the specificities of those data sources. For example, mobile positioning data can supply the number of overnight visits but does not equate precisely to the number of tourists staying at rented accommodation as it is not possible to distinguish hotel visitors from all visitors (although the two correlate very well). This, however, is compensated for by having more complete data on all visitors.

The use of mobile positioning data has the potential to improve several aspects of tourism (and other) statistics, such as timeliness (in some cases up to near real-time), access to statistical information previously not available (new indicators), calibration opportunities for existing data, space and timely resolution and accuracy. Therefore, in the current situation, where existing tourism statistics collection methods cannot be fully replaced, the use of mobile positioning data can on the one hand increase the quality of statistical production and output, but will not necessarily lead to a reduction in the number of processes or in overall workload for NSIs because mobile positioning data will be processed in parallel with the existing system.

Initial implementation and automation is possibly expensive for both MNOs and NSIs, offset over the years by lower costs of maintaining the system. In the end, after the automated processes are in place, the annual work to process mobile positioning data can be drastically less compared to the current methods of gathering tourism statistics. A combination of mobile and traditional methods can also prove to be cost-efficient, e.g. combining mobile data with (smaller-scale) demand surveys may radically reduce survey sample sizes and provide cost-savings.

In addition to official statistics, there are a number of **other additional spheres** that can benefit from the use of mobile positioning including urban planning, transportation planning, monitoring traffic flows, events (concerts, sports, festivals), safety and security (risk analysis based on the number, consistency of people at a specific location at a specific time), transportation origin-destination matrices, tourism and place marketing, epidemiology (geographic spreading of infectious diseases) and monitoring of hot-spots (places of gathering of tourists and residents).

As the compilation of tourism statistics also involves many elements usable in other statistical spheres, joint processes with Balance of Payments (travel item), transport statistics (trips taken by passengers), short-term migration (commuting), long-term migration and population statistics could have a positive effect on the cost-benefit balance since the core data processing processes are the same. Any cost-benefit analysis made by NSIs should, therefore, also take into account the potential usage of mobile positioning data for other fields of statistics, in addition to tourism.

Today, Member States and NSIs often face the same barriers when trying to access and work with mobile phone data. One of the aims of this feasibility study was to reveal, assess and discuss these barriers that are common to most or all European countries.

This study concludes, based on the reactions from the stakeholders and the findings in the different reports, that further work on mobile positioning data as a source for comparable and reliable tourism statistics could benefit from a central framework for NSIs and other interested parties on methodological and legal aspects.

This feasibility study can be considered as a starting point with cross-border relevance towards such a common central framework.

1. Introduction

A key output of the system of tourism statistics is information on tourism flows, i.e. the number of trips taken and the nights spent away by visitors outside their usual environment. Notwithstanding the utmost relevance of economic information (tourism expenditure, monetary flows), the underlying physical flows remain a basic building block of any system that involves tourism statistics. Traditional data sources include surveys that are completed by accommodation establishments (reporting on guests who stayed at their establishments) and surveys filled in by households or individuals (reporting on trips made during a recent reference period). The European legal framework for tourism statistics (Regulation 692/2011) currently in place was developed against the background of these more traditional data sources (Eurostat, 2012).

However, the European Statistical System (ESS) is influenced by changes in the business environment (new data requirements, need to simplify the collection process and reduce the ensuing burden, use of ICT tools), and changes in the ESS business architecture (integration of datasets, reuse of existing data or administrative data) (European Commission 2009). The fast-changing tourism market 'has created new user needs or has changed existing user needs to which the legal framework needs to adapt by introducing new variables or breakdowns, whilst dropping some existing requirements that are no longer essential, and by improving the timeliness of the data,' (Eurostat 2012). Therefore, Eurostat needs to play a role in developing more efficient methods for collecting data, which is also relevant for the field of tourism statistics.

'Mobile positioning data can be a source for monitoring flows of persons, inside or outside their usual environment. In the short or mid-term, it can be an additional source of information for the system of tourism statistics (quick indicators, additional indicators to cover existing gaps — for instance flows of non-residents not staying at rented accommodation, quality checks and calibration of sample surveys), in the long term it could possibly replace part of the existing data collection work. It goes without saying that this would lead to a significant reduction in the reporting burden and to a significant reduction in processing and compilation cost for the Member States' statistical authorities' (Eurostat 2012).

In this context, Eurostat initiated a supranational feasibility study on the use of mobile positioning data for tourism statistics to explore the possibilities — and limits — of this new data source and methodology (Eurostat, 2012). From December 2012 to March 2014, a consortium that consisted of six partners — Positium LBS (Estonia, project lead), Institute for Tourism and Research in Northern Europe (Germany), Statistics Finland (Finland), The French Institute of Science and Technology for Transport, Development and Networks (France), Statistics Estonia (Estonia), and the University of Tartu (Estonia) — conducted this study.

The consortium consisted of specialists in the field of statistics, research, tourism, mobile positioning data processing and location-based services, geographic information systems and transportation.

The methodology explored in the study was the statistical use of location information from technical databases and registries that concern the historical location of mobile devices within the network of mobile network operators (MNOs). The aim of the study was to assess the feasibility of using such data for estimating domestic, outbound and inbound tourism flows and to address those strengths and weaknesses that are related to access, trust, cost and the methodological challenges of using mobile positioning data in tourism statistics.

The consortium concentrated its work on the different aspects of the use of mobile positioning data in tourism statistics and other domains: situation with the use of mobile data; accessibility to the data from legal, technological, financial and business aspects, including possible cost and burden implications; methodological principles of statistical data collection and compilation, including evaluation via different aspects of quality and comparison of the results to those of the existing traditional methods; opportunities and limitations that are inherent in the data source.

During the course of the study, several specific tasks were carried out in order to collect the information needed to fulfil the *ex ante* objectives of this study. The analysis conducted in relation to those projects in which mobile positioning data was used included an overview of the situation in EU Member States, the European Free Trade Association (EFTA) countries and EU candidate countries. A legal analysis based on the legislation of the EU and four Member States (Germany, Estonia, France and Finland) was conducted in order to assess the regulatory factors that are involved in accessing the data from MNOs. A survey was conducted amongst potentially stakeholders (national statistical institutes (NSIs), MNOs, data protection agencies

(DPAs), tourism non-governmental organisations (NGOs), and private businesses) so that an understanding could be reached on interest and experience regarding the use of mobile data in tourism activities and other domains. Several interviews with MNOs, NSIs and DPAs provided a more thorough insight into the specific aspects of legislation, technology and methodology. Though pilot data was requested from several European countries, the main tests in the assessment of methodology and coherence were undertaken with data provided by MNOs in Estonia, the only country for which data was actually accessible during the project.

2. Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

The study involved five different tasks with corresponding reports covering the analysis of the current situation in using mobile positioning data; the aspects of access from legal, technological and business points of view; the methodological aspects of data compilation for tourism statistics; the aspects of coherence for the existing statistical indicators on tourism that are based on the sample mobile positioning data; and the analysis of the opportunities and benefits of using mobile positioning data.

The individual reports are available for download from the **Eurostat website**.

2.1. Stock-taking

Report 1. Stock-taking makes an inventory of all of the research that has been conducted to date, along with applications and experience, in EU Member States, EFTA, candidate countries and around the world, and provides an up-to-date description of the state of the art in using mobile positioning data in research and applications in tourism statistics and related domains. It serves as input for research that has been conducted in the subsequent tasks of the feasibility study.

The report emphasises existing problems and solutions in technology, methodology, and regulations and other aspects of accessibility in terms of mobile positioning data. The information covered in the report is based on the circumstances that existed in summer 2013 regarding the following areas:

- Publicly available information (applications and scientific research);
- Knowledge of the consortium partners;
- Information from surveys and interviews.

The consortium collected and reviewed more than a hundred cases that had included the use of mobile positioning or other ICT-based data that bore any relevance to the study. The above priorities and referenced usage cases were selected because of their direct or indirect pertinence when it came to using mobile data for tourism studies or as a statistical source. A total of 31 significant and representative cases are presented in the report along with an overview of the situation in 36 EU Member States, EFTA countries and EU Candidate States concerning mobile positioning data, research or applications. The full information is available in the detailed report (available on the **Eurostat website**).

The results of the usage case analysis highlight the fact that there is an increasing number of mobile data-based studies, research papers, projects, applications and businesses being created. It can be assumed that in the next 2 to 5 years, such datasets will probably become a common source for studies and statistics in most European countries, and that alongside this process, the data will become available for generating tourism statistics. All successful usage cases were developed step-by-step and success was guaranteed by user persistence.

Society in general is getting used to the idea of implementing sensitive data along with the widespread progression of other privacy-sensitive developments, especially in the social media sphere. We are seeing progress in the data protection sphere, where the last decade has shown a degree of turbulence in relation to the newer aspects of privacy protection, and a good many countries are reviewing their national legislation in order to cope with social issues as well as security threats in the area of data protection. Though tourism statistics neither require nor are intended to track the activities of a specific person, this is something that is closely related to and is perceived as being part of data protection. MNOs are overcoming privacy and business confidentiality-related concerns as they see the appearance of new revenue possibilities, and also the value that can be gained from the internal use of such data. The development of the technology and the principles for processing sensitive data safely and anonymously should allow legislation to adopt and develop the practices for using such data. The analysis of the available research also demonstrated the most valued

strengths of mobile data: relatively simple, quick and cheap collection of data whilst including a large sample. The time (including the level of individual events) and spatial (the accuracy of a mobile cell, potentially GPS precision) accuracies are also better than previously. The problems related to the dataset include a lack of qualitative information on the user, such as the purpose of the visits and the means of transport; as well as privacy issues and the problems that can accompany the processing of large amounts of data.

Most of the time, access to data results in long-term, trust building cooperation between all of the parties involved, in which projects grow from small-scale testing projects to wider collaborations. This has led to a step-by-step process of obtaining larger amounts of data with better quality.

A second group of companies have direct business solutions focused on providing technology for MNOs. New business projects are started in order to promote the operator through services beneficial to society. There are also cases with clear profit models.

The third group of cases involved in obtaining data consists of short-term projects, such as Real-Time Rome or sporting competitions in Milan. MNOs agree to reveal 'a bit' of their data for certain events or projects for the purpose of publicity or for testing potential business opportunities. Several large-scale projects have started from such small ventures.

Based on findings from the usage cases that have already been examined, it can be concluded that mobile data is being used increasingly in a number of different fields. Most of the active usage is within academia, with some already established applications on a state level (e.g. E-112, Bank of Estonia). There are a few direct usage cases and examples that cover the use of mobile positioning data in generating tourism statistics. The preliminary results from the Task 2 questionnaire show, however, that many of the statistical bodies have already considered the use of such data or have even contacted MNOs with data requests. It was discovered that a major problem for statistical bodies was not methodological but concerned access to data, privacy concerns, and the relatively high 'entry cost' of using new ICT-based data sources.

Business-oriented projects that are the focus of the mobile operators concentrate on geomarketing; non-MNO models focus on providing technology and/or brokerage of the results data for usage in transportation, traffic, urban studies, regional development, and tourism applications.

2.2. Feasibility of access

Report 2. Feasibility of Access assesses the potential opportunities and obstacles in terms of gaining access to passive mobile positioning data [3] from mobile network operators. The main focus is on access to the data in order to produce official tourism statistics for the NSIs, but other usages are also considered. The report concentrates on regulatory, business and technological barriers along with practical access to the data. The knowledge for the current report is obtained from the experience of the consortium from the past projects, legal analysis of the subject, the practical process of accessing the pilot data from a number of MNOs or other organisations that have acquired the data, and the experience of others gathered from the online survey that involves relevant organisations and face-to-face interviews with experts.

All of the stakeholders including NSIs, tourism promotion organisations, research institutes and MNOs seem to realise the potential of utilising this data source but see many obstacles in terms of privacy protection and legislation, user adaptation (the switch to new data sources from traditional practices), the representative nature of the data, and various methods for translating mobile data so that it represents the 'real world'.

2.2.1. Regulations and Privacy Protection

A legal analysis was conducted, which covered regulation and privacy protection aspects and which consisted of an examination of the legislation at EU and national levels (in Germany, Estonia, France and Finland) relevant to the subject of the study. The aim of the analysis was to assess the current situation and to provide insight in relation to current and future options when it comes to accessing the data.

Legal restrictions are considered to be the most important barrier when it comes to accessing the data. The legal acts do not specifically concentrate on the use of mobile positioning data in statistics. The relevant areas of the legislation on the EU level include:

[3] **There are two main technological methods for locating mobile phones from the infrastructure of MNOs: active and passive mobile positioning. Active positioning locates the owner of the phone in real-time and requires the owner's consent. Passive mobile positioning is extracting data representing information about historical locations of the phones from the log files of MNOs. This latter method allows for a longitudinal view of all subscribers and was the focus of this study.**

- the Data Protection Directive (Directive 1995/46/EC and its successor, the General Data Protection Regulation);
- the Electronic Privacy Directive (Directive 2002/58/EC);
- the Data Retention Directive [4] (Directive 2006/24/EC);
- the European Statistics Regulation (Regulation 223/2009/EC);
- the European Statistics Regulation on tourism statistics (Regulation 692/2011/EU);
- and the opinions of Article 29 Data Protection Working Party [5].

These legal acts stipulate the following main general principles for the use of mobile positioning data:

- Directly or indirectly identifiable mobile positioning data can be used and processed for statistics if the following is true:
  - (a) The subscriber has given their consent for their data to be processed for such specific purpose; or
  - (b) The official statistical organisation (the NSI) needs the data in order to perform an official task imposed on them by the law.
- Fully anonymous mobile positioning data can be processed and used without restriction given that subscribers cannot be directly or indirectly identifiable in any given processing stage.

The two options set out the clear limitations on the methodology used for processing the data in theory. The full methodological opportunities open up when it comes to the use of the longitudinal data, which makes it possible to calculate the crucial aspects of tourism activities (and other domains), such as defining the usual environment, the duration of visits, movement paths during the trips, etc. However, the longitudinal data on its own is a factor that can allow the identification of individual persons. The anonymisation methods currently being used exclude any options available in relation to long-time analysis (e.g. the aggregation of the initial data before it is processed), and therefore seriously limit the available options in relation to methodology. There is a need for research programmes to develop the process of anonymisation so that it is able to preserve the longevity of the data, as no conclusive methods have been produced so far.

An alternative interpretation can be formulated, which decrees that, because the end result of processing is, by itself, anonymous (involving an aggregated number, duration, and the travel patterns for collective trips), the processing of personal data for such a purpose can be interpreted as not being illegal. However, such an interpretation can be considered as a grey area when it concerns privacy protection legislation as the legislation states that any processing of personal data, irrespective of the results, falls under the legal constraints of data protection legislation.

The practical implementations of these principles are left for the Member States to determine. In the course of the study, the corresponding legislations in Germany, Estonia, France and Finland were analysed. Although the main principles are taken from EU directives, the practical implementation varies and can be interpreted differently. These differences in national legislation concerning the use of mobile data make interpretation and understanding even more obscure. The differences can be, for example, in the scope of the national statistics acts that govern the area of statistical data sources for national official statistics. The clause that concerns the use of personal data (which is either directly or indirectly identifiable) for the purpose of historical or scientific research (Article 11/2 of the Directive 1995/46/EC) can also be interpreted differently as there is no clear definition of 'historical or scientific research'. This results in a situation in some countries in which mobile data can be accessed and used more easily and in wider areas, while it is almost impossible or very limited in others.

Bearing in mind the current European legislative framework, it can be stated that the use of mobile data could be implemented differently in each Member State, which results in possibly different methodologies and outcomes against the harmonisation of results desired at European

---

[4] **The European Court of Justice has declared the Data Retention Directive to be invalid in its judgement of 8 April 2014, i.e. after the date of this Report. It remains to be seen when and to what extent the member states' relevant local laws that were enacted under the Data Retention Directive will be reviewed. Link: http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-04/cp140054en.pdf**

[5] **On 10 April 2014, i.e. after the completion of this Report, the Article 29 Data Protection Working Party adopted their Opinion 05/2014 on anonymisation techniques analysing the effectiveness and limits of existing anonymisation techniques and giving guidance on how to handle these in light of securing the privacy of individuals. Link: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf**

level. The most evident option is to amend or enable national statistics acts so that they compel MNOs to provide the data for NSIs either by extracting and transmitting the raw data to NSIs or by processing the data within MNOs and transmitting the resulting data to the NSIs. Such an option is 'limited' to the definition of official statistics, and therefore possibly excludes new indicators (some of the main advantages of the methodology) simply because they are not required by official statistics.

The current European tourism statistics regulation concentrates on data sources that are used today, which to some extent results in statistical indicators that are driven by the specific data sources that enable to compile these statistics (e.g. accommodation statistics) rather than being driven by the full set of user needs. During the course of the study, several respondents from the survey proposed the development of a central European framework for NSIs in order to enable them the access to mobile positioning data and simplify the process of obtaining data. This would mean either the inclusion of such data sources, the corresponding statistical indicators and classes in the list of official and mandatory statistics or the drafting of more specific and clearer articles within data protection and electronic communications legislation that would make it easier for MNOs to provide the data for NSIs or other users based on the assumption that the purpose behind its processing is purely the collection of statistical information and not any intrusion into the privacy of the individuals concerned. However, this is a lengthy process at best, and it requires in-depth legislative work.

2.2.2. Business-related Barriers

Public Opinion

Besides the legal barriers, the use of mobile positioning data represents a challenge from the point of view of public opinion. The data processing not only has to be legal and the methods being used not incorporate any intrusion into the privacy of the subscribers, but this has to be presented as being appropriate to the general public. Mobile data is highly sensitive and can present risks in terms of data misuse. The current study does not focus on the opportunities in terms of surveillance activities by law enforcement agencies and does not assess the definition of data misuse, but obviously there are several ways in which the use of the data can be perceived by people as being wrongful. The main idea behind the use of the data for collective statistical purposes, while not focusing on the individual profiling of the subscriber, is by itself understandable and noble; however, it can be perceived by the public as a means for governments or mobile operators to 'track people'.

Data revealing the location of mobile phones has been the central argument of a perceived misuse on several occasions. The lessons from those precedents show that for a successful use of mobile positioning data, the key players (NSIs, Data Protection Agencies, and MNOs) have to be very serious in acknowledging to the general public that such use of the data does not involve the tracking of individuals and that the purpose is to compile collective, anonymous statistics concerning the phenomena of tourism activities (or another relevant domain). Involvement of NGOs and journalists, publication of the results and explanation of the processes are rather crucial as it has been established that fears of tracking, be it rational or irrational, are driven by a fear of the unknown, a sense of being watched but without knowing by whom or for what purpose.

Burden for MNOs against Benefits

MNOs are interested in discussing the potential arising from the data that they have at their disposal. However, several aspects of data acquisition directly affect the internal questions that MNOs raise: starting with the MNO burden (which involves technological implementation, human and technical resources, and maintenance of the system); business confidentiality aspects that are important for MNOs in order for them to be able to preserve their competitive advantage; financial or other benefits that MNOs can potentially gain but which also include the aforementioned legal aspects and the effects of public opinion on the number of clients.

MNOs are cautiously optimistic, which means they also see potential financial benefits from using their assets (in terms of positioning data) and creating new revenue streams. However, if the data is obtained through enforced legislation that only sets up obligations and burden without providing financial or any other form of compensation, MNO's enthusiasm decreases. Instead, MNOs expect to benefit from providing the data for the state either through some form of financial compensation or by being allowed to use the data internally or externally for other purposes. For example, improved tourism statistics or population statistics can potentially be beneficial for MNOs in that they can improve their analysis of penetration rates among roaming customers or the geographical market share (comparing the number of the roaming subscribers with the total number of tourists or the number of subscribers in specific geographical regions to the number of residents in the area).

Financial interests were one of the most frequently mentioned reasons after legislation for not providing data to interested users from the responses gained through the survey and interviews, which was further confirmed by the consortium's efforts to access pilot data for the current study. MNOs have to take into consideration the human and technological resources required to provide the data. In the case of MNOs, these might be expressed in substantial financial figures.

The cost of the system that retrieves and processes the data within MNOs consists of implementation (i.e. the initial investment required to set up the system) and maintenance costs (i.e. the price to be paid for keeping the system working), and depends on several variables such as the size of the MNO (the volume of data that has to be processed), the allocation of processing resources (i.e. whether MNOs only have to extract and deliver the raw initial data to the external processing agency or are required to implement the full processing chain), the number of processes to be conducted by the MNO (i.e. the forms taken by the data: inbound and/or domestic and/or outbound; geographical probability calculations for usual environment, etc.), the maximum allowed latency levels (i.e. the maximum allowed processing time from the point at which the initial raw data is extracted to the delivery of the data to the NSI), and finally external implications (e.g. licensing external technology, outsourcing costs, and the costs involved in using internal resources).

The total cost for such a system (involving automation of the full processing chain and for all forms of tourism activity) within one MNO with 10 million subscribers and an obligation (latency/timeliness) to deliver the results every 15 days can be roughly estimated at EUR 550 000 for implementation and EUR 160 000 per year for maintenance. Please bear in mind that the costs presented here are very rough estimates — giving a possible order of magnitude — because only MNOs themselves can provide the actual costs, which will be based on their internal calculations. The burden on each MNO for continuously providing data may vary greatly depending on the size and internal system complexity of the MNO in question. In addition to the implementation and maintenance costs for the MNO, there are quite naturally costs that will be incurred by the receiving party (i.e. the NSI).

Obviously the internal costs for the MNO depend on the size of the MNO and the required latency (see **Figure 1**). The cost of shortening latency time might increase exponentially as it nears real-time when substantial resources are needed to process the data, in addition to the fact that maintenance requires the constant monitoring and attention of specialists.

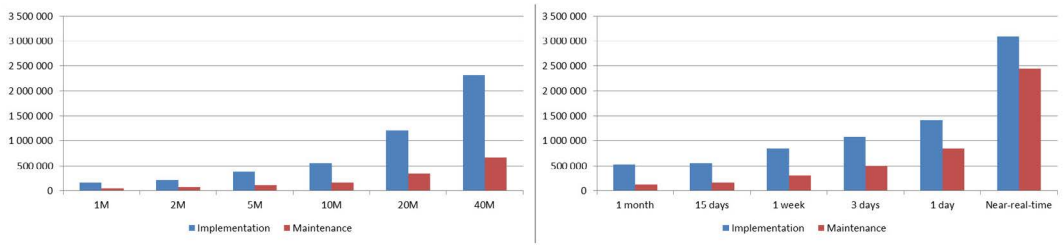

Figure 1. Estimated cost of the implementation and yearly maintenance of the system for MNOs for the production of mobile positioning-based statistical indicators based on the size of the MNO (in millions of subscribers) and the latency (timeliness) of the processes (in EUR). Left: Latency 15 days. Right: MNO with 10 million subscribers.

The differentiation between the financial burden of MNOs as data providers and the NSI can be illustrated as follows: if MNOs have to process the data and deliver the aggregated results to the NSI (see **Figure 1**), then the cost of the overall system is higher (more processing systems for multiple MNOs, a simple system for the NSI); if the MNOs are only to extract and transmit the raw positioning data to the NSI, then the cost for MNOs is lower, but the NSI has to invest more in their data processing system.

Confidentiality Issues with Business Secrets

Confidential business information essentially means strategically important information about the activity of MNOs. Access to such information by other MNOs could potentially harm the competitive advantage of the MNO in question. MNOs consider the loss of any of their own sensitive business information to competitors to be a far greater loss than any gain they might be able to make by learning the confidential business information of other MNOs. MNOs are primarily concerned about the possibility of their competitors acquiring the following information:

- Their number of subscribers (both domestic and roaming);
- The number of their service activities (calls, messaging, data) in the network;

- Any information on the constitution of their subscribers (the number of pre-paid versus post-paid cards, the socio-demographic information of their subscribers, the number of subscribers in various foreign countries, etc.);
- The number and locations of their network antennae (the release of this information might also be prohibited by law in some countries due to the terror threat to that country's vital telecommunications infrastructure);
- Any financial information and strategic plans of MNOs;
- Technological capabilities and information on infrastructure and systems.

This issue is clearly an important one as it was repeatedly mentioned by MNOs, but it is one that can be resolved by determining procedures that do not allow such information to be exposed to competitors.

Patents

The relevance of understanding patents and intellectual property (IP) rights for this feasibility study is that it provides an understanding of ongoing technological developments and their direction as they take place in the European and global market place. The collection and processing of data for tourism statistical purposes — as explained in previous chapters — is technology-intensive and sees the application of a range of data processing methodologies and techniques.

Because the actual relevance and potential infringement threat is fairly difficult to establish, it is not within the scope of the current report to name those patents that are barriers in generating tourism statistics. The list that has been presented in the report is merely an indication of the patents that may or may not be relevant to the technology that will be used to generate tourism statistics. However, such a list might also present opportunities to involve the technologies mentioned as their use might improve the quality of the data or simplify some processes.

2.2.3. Technological Aspects

The technological and methodological aspects of processing mobile positioning data are tightly linked. Report 2 thoroughly discusses the technological areas of this subject with the aim of helping readers understand the nature of the initial data that is in possession of MNOs, starting from logical data movement flows within a single MNO's core network systems and progressing up to the initiation of data processing specifically for the purposes of tourism statistics. This discussion includes the processing and preparation of the data for the formation of a framework and for data compilation requirements, including the identification of the home country (in inbound roaming data), handling subscriber identity codes, the geographical and time properties that are included in the data, any available additional attributes (e.g. socio-demographic attributes for subscribers or the technical attributes of the event), the removal of non-human devices, the use of a blacklist and sampling, and the formatting and preparation of the data for further tourism-specific processing.

Where technical access and data processing are concerned, the main questions are connected to the specific data sources — there exists a wide range of databases and registries that can be used; however, they often differ between MNOs depending on the system architecture being used and on their technical ability to be able to store different types of data.

Three main types of data can be distinguished, based on their origin within the MNO's systems, where such data is required for the compilation of tourism statistics:

(a) (any kind of) event data (metadata) that covers subscriber activities and which is included in the MNO data stream;

(b) geographical cellular (network) referencing data;

(c) attribute data for subscribers (e.g. demographic information taken from the customer database).

Event data can be divided into internal and external network events and, furthermore, it can be broken down in the same way as for the various forms of tourism statistics:

(a) MNO internal events:
   1. inbound roaming;
   2. domestic;

(b) MNO external events:
   1. outbound roaming;

The most common source for mobile data, as well as being the easiest to access, is Call Detail Records (CDR), with it being possible to take data from inbound and outbound roaming and domestic datasets as this data is held in storage and is rather easily accessible by the MNOs, although at the same time this is the most tenuous (in terms of the number of

records/events per subscriber per day). The CDRs represent the active usage of mobile devices — covering incoming and outgoing calls and SMS text messaging. The biggest problems with CDRs are the frequency and the regularity of the records as they are based on the usage pattern of the subscriber. The average number of CDRs for tourists is approximately four events per subscriber per day — meaning that there is an average of four location facts for a phone for every single day. This is sufficient for some areas, but it sets limits upon domains in which better temporal accuracy is required (e.g. hourly statistics on a very small geographical space). Alternative data sources such as Data Detail Records (DDR), location updates, or others can include up to several hundred location events per user per day; however, such data is not often stored by MNOs.

The geographical locations of the events are retrieved from the reference between the events and the geographical attributes of the network antennae. By default the location of the antenna is considered as the location of the event. This guarantees a rather low level of geographical accuracy for specific events. However, unless there is a more advanced technology available to use - either the GPS signals or additional accuracy improving technical attributes (e.g. distance from the antennae) - it is not possible to precisely define the location of the mobile device. However, for data that includes thousands of events per day, a probabilistic distribution of events in space using land coverage information (covering the road network, residential areas, forests, and fields) could result in a better geographical representation of events (see **Figure 2**).

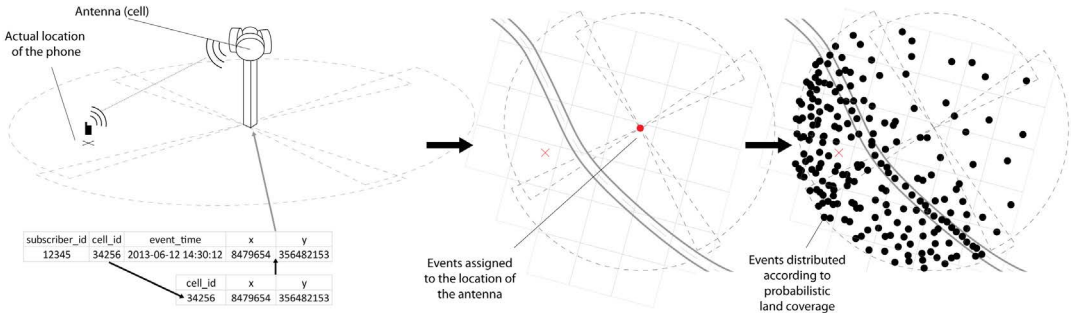

Figure 2. An illustration showing event coordinate mapping using single point antenna coordinates and probabilistic distribution.

The size of the data block, the number of records it covers, and the processing complexity it creates all require a sophisticated data processing system that can roughly be divided into two options (see **Figure 3**):

1. Data is extracted and processed within the MNOs, and the resulting statistical indicators are transmitted to the NSI, where the results from several MNOs are combined to create the final statistical indicators;
2. Data is extracted by MNOs and transmitted to the NSI, where the processing is carried out in order to produce the final indicators.

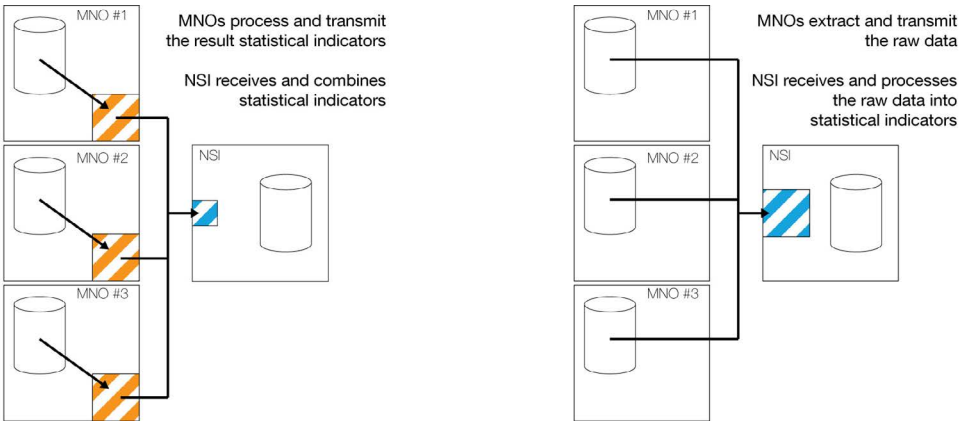

Figure 3. Two options for processing the mobile positioning data.

The cost and distribution of the burden are different for both scenarios. In addition, there is no clear preference as both these options have their clear benefits and disadvantages. The main technological challenge lies in the ability of the system to carry out periodical complex processes involving large data records within the designated timeframe with the chance of being able to recalculate the results in case any error occurs.

2.3. Feasibility of Use: Methodological Issues

Mobile positioning data describes mobile phone usage and as such is not directly meant to be used in official statistics. Therefore the processing steps are quite different from the processes used for traditional survey data. *Report 3a Feasibility of use: methodological issues* describes the methodology of tourism statistics that cover processes from obtaining raw data until the results are published. The described methodology is evaluated by using different quality aspects. In addition to tourism statistics, the possible relevance of mobile positioning data for other fields of statistics is addressed and an assessment of how well-suited the described methodology is for the usage of mobile positioning data in other fields is provided.

2.3.1. Data Processing Methodology

The technological section of *Report 2* concentrates upon initial data sources, the technologies that can be used to extract the data from MNOs, and the preparation of the data for further processing, specifically for tourism statistics. In *Report 3a*, the steps following data preparation by MNOs are listed and described.

During the processing of mobile positioning data, there are several important steps that enable generating the tourism statistics, such as the identification of usual environment and the country of residence, the duration of stay in specific place, differentiation between same-day and overnight visits, etc.

Depending upon the availability of the data and upon technological availability, all of those trips that are taken within the framework of the requirements can be analysed where they can be seen to correspond to the situation shown in the census or, alternatively, a subset of observations within the framework could be selected. The sample sizes can be substantially larger in this situation when compared to traditional sample surveys, as the cost and burden of data collection is much less driven by the number of observations in the sample. The sample size can be determined from available technological capabilities and disclosure rules. The aspects of cost and burden are discussed in the respective chapter of Report 4.

The methodology contains the following sections: the additional preparation of event data, frame formation, data compilation and estimation. The initial data extracted and prepared by MNOs is based upon network events that specify a specific subscriber's presence in time and space. Additional preparation may include geographical referencing, the elimination of non-human-operated mobile devices, checking the time and area coverage of the data, dealing with missing values, etc. After the data has been prepared by MNOs, the following processing steps are set out:

- Frame formation:
  - ◊ Applying trip identification algorithms — identifying each subscriber's individual trip to the destination (country of residence, foreign country) in question with the start and end times for each trip;
  - ◊ Identifying the population of interest (distinguishing tourism activities from non-tourism activities):
    - * Defining roaming subscribers not actually crossing the border and entering the country (inbound, outbound);
    - * Defining residents (inbound, outbound);
    - * Defining the place of residence and the usual environment (domestic);
    - * Identifying country-wide transit trips (inbound);
    - * Identifying destination and transit countries (outbound).
- Data compilation:
  - ◊ Spatial granulation (visits at the smallest administrative level for inbound);
  - ◊ Defining variables (number of visits, duration of trips, classification, etc.).
- Estimation (from an MNO-specific sample to the whole population of interest) contains:
  - ◊ Time and space aggregation of the data (day, week, month, quarter/grid-based (one km2), LAU-2, LAU-1, country);
  - ◊ Combining data from various MNOs and computing final statistical indicators.

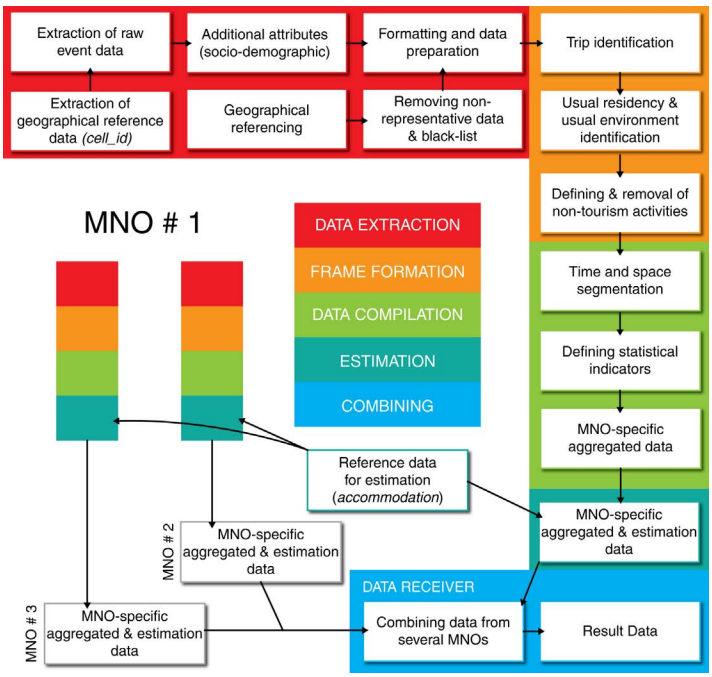

Figure 4. Data processing steps.

Figure 4 represents the general steps taken from processing the raw data to the final aggregated and estimated results.

The specific of methodological aspects might depend upon the characteristics and origin of the data (time and space frequency, the geographical accuracy of the events, and the available attributes of the events or the subscribers). The processes described in Report 3a (trip calculation, identification of usual environment, non-residents, transit trips, duration of stays, etc.) assume that longitudinal calculations for single subscribers are possible.

Essential procedures carried out during the frame formation process include the identification of the county of residence and the usual environment of the subscribers in order to be able to define tourism trips (i.e. trips outside the usual environment). These calculations require the historical time series (longevity) and interconnectivity between the different data forms (same subscriber ID in domestic and outbound data) for the subscribers in order to be able to define frequently visited places and countries.

This methodology description in *Report 3a* could be taken as step-by-step guidelines on how to produce tourism statistics as it is rather detailed but at the same time general enough for broad use. There is an assumption that several variables relevant for the forming of new variables exist in the dataset and that the activities of anonymous subscribers can be followed over a longer period of time to establish their residency and/or usual environment. The assumption regarding the availability of longitudinal data is crucial for the production of tourism statistics. Algorithms identifying the usual environment and country of usual residence rely upon the availability of past data. If the available data describes only a short period of time then the issue of processing errors arises and, as simulations show, the data quality can be too low to produce reliable results. Such limited data can be used as comparison indicators in some unofficial domains (e.g. the number of unique foreign subscribers on the site of attraction or concert) and for relative comparison.

2.3.2. Quality of the Methodology

The methodology is evaluated with regard to validity, accuracy, and comparability. Other evaluation aspects are covered in different reports for this study, such as accessibility (*Report 2*), coherence (*Report 3b*), timeliness (*Report 4*) and costs (*Report 4*).

The standard quality aspects are reviewed here to assess and improve the quality of the output. Although the same aspects are used to evaluate the estimates based upon survey data, there are differences in the relevance of these aspects to the output. For example, sampling errors and non-response errors are often measured and evaluated in traditional sample surveys as these are usually the main error sources.

There are several inherent data limitations that impact the quality of the methodology. The shortcomings that are inherent in mobile data include the lack of information on expenditure, the purpose of the trip, the method of transportation and the overall qualitative aspects

of the tourism activity. It should be noted that the quality of the final outcome relies heavily on the availability of external information such as accommodation statistics, transport statistics, information about the market share for mobile operators and their subscribers and other information used for estimation. The quality of this information should be assessed before using it as auxiliary information in the estimation models. Overall, the estimation part of the described methodology is very general allowing a frequentist or Bayesian framework to be applied.

Many quality issues that are presented by using mobile positioning data in official statistics are similar to those faced when starting to use a new administrative data source — differences in the concepts and definitions, coverage of the population frame, the representative nature of the sample, etc. In the case of administrative data or in mobile positioning data, NSIs do not have any influence over content, so an assessment of the validity is mainly qualitative (consisting of a description of the discrepancies between what is measured and what should be measured).

When comparing the definition for the target populations (on its participation in tourism activities and on the characteristics of tourism trips) against the population frame, there is one discrepancy that is immediately clear. Within the frame are all of those subscribers who used their mobile phone for calling or texting whilst the population shows all of those individuals who reside in the country. This leads to a large number of various coverage problems, the complete list of which is discussed in *Report 3a*.

The accuracy is the most problematic quality aspect for this type of data, especially in terms of coverage issues. It is problematic because there are many components that contribute to the coverage bias and assessing all of them, separately or together, is a very complex task. There is no one method available at the moment that will allow easy estimation of the different biases. For several quality issues, quantitative results are given in the report based upon Estonian mobile positioning data to describe and illustrate the problem.

The quality assessment for assuring the comparability should be carried out when changes in methodology occur (e.g. due to the number and structure of MNOs), just as it should be carried out in traditional surveys. In addition, it is important to be ready to update the methodology if changes in the telecommunications technology or in the data structure occur.

There are many contributors to the coverage bias, but due to the co-effect some bias components cancel each other out (over-coverage versus under-coverage), some contribute very little, and some may contribute a lot. Many problems, however, are inherent in mobile positioning data and therefore cannot be avoided. Furthermore, their total effect, i.e. the total size of the coverage bias of an estimate of interest, needs to be evaluated or bias-corrected estimates need to be computed. One should note that, for example, a fairly large percentage of people who do not possess a phone does not automatically lead to large coverage bias. A bias occurs when those people who do not have a phone travel a lot. Unfortunately, there is no information on the relationship between owning and using a mobile phone and a person's travelling habits. It's worth recalling that coverage and other bias also applies to existing more traditional data collection methods (e.g. CATI interviews and dropping penetration rates of fixed lines).

Similarly to traditional surveys, comparability can suffer for statistics based upon mobile positioning data when changes in underlying legislation occur or when bigger changes in the methodology or technology are introduced. The impact of these changes can be at best minimal or non-existent but they can also lead to a break in the series depending upon the nature of the change. Possible technological changes and their impacts are discussed in *Report 2 Section 4.6: Methodological Changes* (changing the frame, processing algorithms, and estimation methods) need to be carried out by taking into account the effect experienced by all quality dimensions and by avoiding changes that affect quality.

From the point of view of tourism statistics and other domains in which passive mobile positioning data can be used, it is important to be able to use the data continuously over a longer period of time for the sake of comparability over time and between regions. There are three main causes for this continuity to be altered. In all cases there can be positive and negative effects that might or might not have an effect on the quality of the data:

- Major global shift in mobile technology;
- Changes of the characteristics of the data;
- Administrative changes (e.g. changed number of providing MNOs).

It is difficult to foresee the final effect of the changes on the results when the changes have just been made. In some cases changes might not affect the outcomes at all if the methodology and the estimations are adjusted properly. Ideally, the inclusion of all MNOs in the country should show an insignificant increase in

quality levels when compared to a single MNO with good coverage that is representative of the whole population and an adequately implemented methodology and estimations. Each change should be well assessed before any adjustment is made.

If all providing MNOs cease to provide data, then there is not much to do as historical data cannot be used to project future trends. Therefore, the legal framework for using the data should minimise the chances of MNOs bailing out.

Some technological changes in the essence of the data (characteristics, type of data, etc.) might require the recalculation of historical data. In such a case the recalculation can only be carried out using the stored initial data and if the effect on the results is major, then it will not be possible to compare the old and new results. Ideally, any changes in the data source should not result in a change of methodology. If changes in the methodology are required, then such a change should aim to produce the same results as with previous methodologies.

The flexibility to introduce the changes depends on the configuration and the set-up of the system. From the point of view of the allocation of the technology the changes are more easily adopted when the processing resides outside of the MNOs and MNOs only extract, prepare and transfer the data to the processing party (e.g. the NSI). In that case, only changes in the extraction processes are needed from the MNO side and it will not be necessary to implement complicated data processing modifications across all MNOs.

Due to the nature of passive mobile positioning data, the quality of the estimates based upon this data source depends on changes in the telecommunications market, e.g. the cost of calls and text messages and the way in which individuals use their mobile phones. Mobile phone technology has developed very rapidly and people use mobile phones for much more than simply calling and texting. It can be seen that the increased level of options available (e.g. mobile broadband) will change people's phone usage habits, resulting in a change of the data content (e.g. less calls and messaging events and more data on internet usage — DDR). As a consequence, the use of advanced data (in addition to CDRs) is called for in order to maintain and improve the quality of the data.

Concerning the quality issues, there are several of them to take into account when processing mobile positioning data: the differences in the concepts and definitions when compared to the current official statistics; the over-coverage and under-coverage of the population frame; sampling issues; measurement and processing issues; and the comparability questions. From a statistical point of view, the main concern is the selection bias. Indeed, even if its diffusion is high all over Europe, not everybody uses a mobile phone, and many people have more than one mobile device. As the data cannot be easily obtained from all MNOs, any change in the structure of this sector would introduce a bias in the measurement of evolutions. However, our tests have shown a good fit between mobile positioning data and conventional monthly time-series on examples.

2.3.3. Relevance for Other Fields of Statistics

Although tourism statistics were the main focus of this study, several other fields of statistics may benefit from this source if they use the same or similar definitions and joint processing schemes as in tourism statistics. Based on the analysis carried out, the following areas of official statistics might benefit if they were implemented as a single system based on mobile positioning data: Balance of Payments travel item, Passenger Transport, Population, Migration and Commuting Statistics (see **Figure 5**).

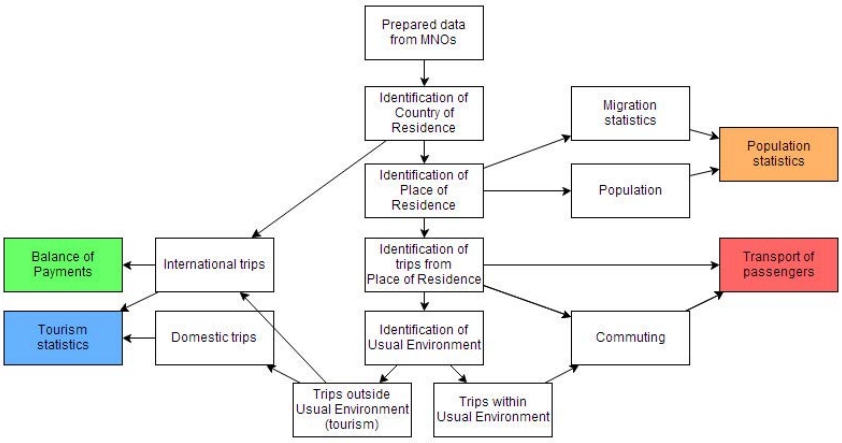

Figure 5. A simplified scheme of joint data processing for tourism activities and other statistical domains.

In economy and finance statistics, mostly the Balance of Payments (BoP) is relevant for assessing synergies with tourism statistics. Statistically, 'tourism' is a subset of 'travel' and consequently, (tourism) 'visitors' is a subset of 'travellers'. Moreover, the issues for tourism activities, namely the problem to accurately delineate the usual environment, do not apply to the travel item in the Balance of Payments. Methodologically, during the processing of data, several objectives can be achieved depending on the system's setup. For example, although domestic tourism concentrates on travel outside the usual environment, the same process can be extended to identify any trips taken within the usual environment.

For calibrating transport demand and organising transport supply, it is very important to have accurate estimates of origin-destination matrices. However, it is quite difficult and very costly to obtain these matrices through conventional survey methods. Therefore, MNOs can provide less costly and much more accurate matrices from mobile positioning data. However, the data will not show the mode of transport or the purpose of the trip.

In population statistics, the spatial distribution of population (living, working) and mobility aspects such as commuting will be relevant. As opposed to census-based statistics, mobile positioning data will always lack accuracy and will not offer the currently required level of detail in terms of socioeconomic breakdowns. Nevertheless, the data is timely and can provide overall indications concerning commuting, migration and internal migration information.

The implementation of a system of statistics production based on mobile positioning data is rather expensive; however, if the system is implemented for several domains (tourism activities including BoP, transportation and population), the additional costs for adding processing components is relatively lower.

2.4. Feasibility of Use: Coherence

The focus of *Report 3b. Feasibility of Use: Coherence* is on carrying out a quantitative and qualitative comparison of mobile-phone-based tourism statistics with reference statistics that include official tourism statistics and other available indicators related to tourism activities. The qualitative element of the report is to describe the reasons for deviations in these data sources. The report does not repeat the methodological issues described in *Report 3a* but rather quantifies the total impact of the methodology used to produce statistics on tourism flows based upon mobile positioning data.

The analysis is carried out with mobile phone based statistics from Estonia ([6]). These statistics are compared against the official tourism statistics made available by Eurostat and against other indicators related to tourism statistics.

The key questions addressed in the report include the following:

- How exhaustively do positioning-based statistics cover physical tourism flows as measured by the reference statistics?
- How well does positioning-based data support a breakdown into the various components of tourism activities: inbound, outbound, domestic tourism and/or same-day trips, paid accommodation, and free accommodation?
- How does positioning-based data compare to existing statistical indicators?
- What are the reasons behind the deviations in existing statistical indicators? Are these deviations strengths or weaknesses of mobile positioning data?
- Does positioning-based data support previously unavailable levels of detail such as geography, time, nationality, etc.?
- Are there any previously unavailable metrics that can be compiled using positioning-based statistics?

Tourism statistics from mobile positioning data were provided by Positium, based upon datasets from two Estonian mobile network operators. The data includes Estonian inbound, outbound and domestic tourism between 2008 and 2012.

The reference statistics were provided mainly by Statistics Finland and Statistics Estonia. Reference data, available on the Eurostat website, on the outbound trips of Estonians to EU Member States was also used.

The mirror statistics method is used to compare two metrics that describe the same scope of observation. An example of mirror statistics is the number of Finnish tourists in the Estonian positioning-based inbound statistics when compared to the number of Finnish tourists in the official Finnish outbound tourism statistics. These comparisons are visually plotted and are analysed using statistical indicators that measure

([6]) **Data from other countries was not available for the study due to access problems.**

both coverage (CC2-value ([7])) and consistency (r-value ([8])). An r-value close to one indicates a high correlation with the reference statistics. Similarly, a CC2 value close to zero indicates a close match in the absolute number of trips.

A sample of these comparisons provides an overview of the consistency of mobile positioning data compared to various mirror statistics, as visualised in **Figure 6**.

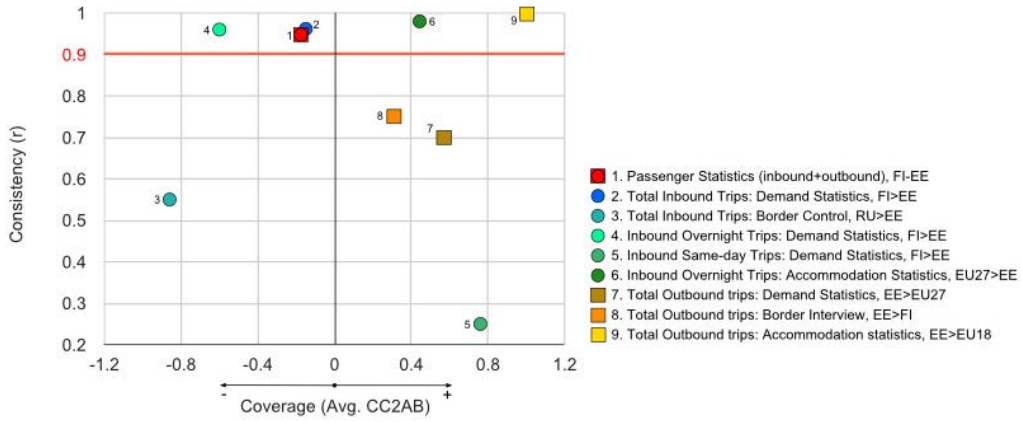

Figure 6. An overview of the two observed coherence indicators (consistency and coverage) compared against various mirror statistics (Demand, Supply, Border and Passenger Statistics).

A few relevant examples from the study are presented here to illustrate the different aspects of the coherence.

Ferry passengers travelling between Estonia and Finland and the combined ferry and passenger numbers between Estonia and Sweden provided good testing corridors for analysing the total volume of inbound and outbound between Estonia and Finland/Sweden. The strength of these passenger statistics is that it is an administrative data source that includes a vast majority of the total number of tourism flows between the countries. However, a drawback is that unknown number of other nationalities and in-transit passengers are on board as well, which cannot be compared with mobile-based statistics. Based upon the 2009–12 monthly time series it can be concluded that MDP provides a very consistent estimate of the total number of inbound and outbound trips (see **Figure 7**). Mobile positioning data underestimates real passenger volumes as reflected by reference statistics due to the other nationalities and transit passengers on board.

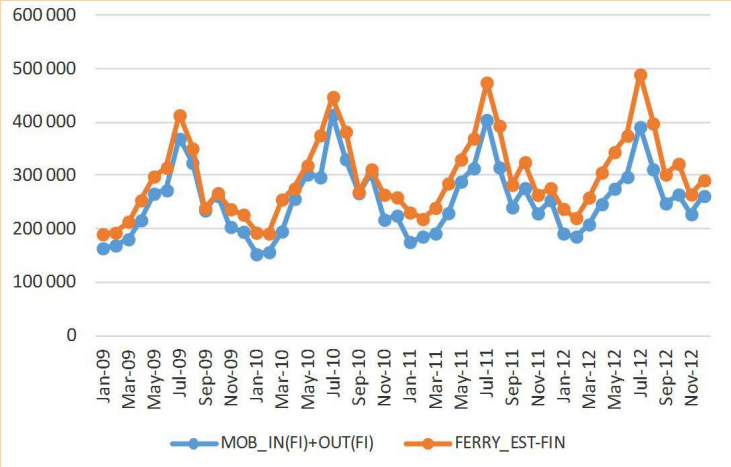

Figure 7. Monthly trips inbound (Finns to Estonia) + outbound (Estonians to Finland) in mobile positioning data when compared to the number of ferry passengers sailing between Finland and Estonia (*Source:* Positium, Finnish Transport Agency).

([7]) **CC2: asymmetry for mirror flows statistics — coefficient. The absolute difference between inbound and outbound flows to and from a pair of countries divided by the average of these two values.**

([8]) **R-value: The indicator that was chosen for measuring consistency over time is the Pearson product-moment correlation coefficient or, in short, Pearson's r. The indicator measures the linear correlation between the two variables X and Y.**

When compared to demand statistics, mobile positioning data also provides a consistent estimate of the total number of trips of Finns to Estonia when compared to Finnish demand statistics. In 2012 mobile positioning data slightly underestimates the number of trips when compared to demand statistics (see **Figure 8**). The main potential reason for this underestimation is that some Finnish tourists to Estonia either do not have any mobile phone or do not use them during the trip.

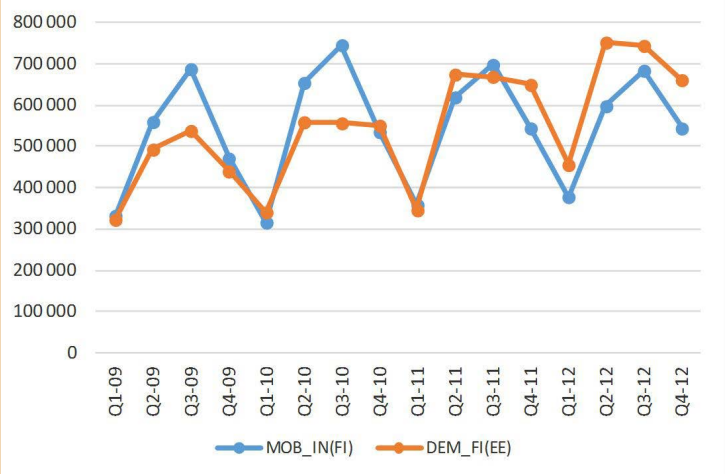

Figure 8. Quarterly inbound trips (Finns to Estonia) in mobile positioning data when compared to the total number of trips to Estonia in Finnish Demand Statistics (*Source:* Positium, Statistics Finland).

The consistency of mobile positioning data against outbound trips in Estonian demand statistics is only moderate. The higher number of trips in mobile positioning data when compared to demand statistics implies that mobile positioning data contains many Estonian outbound trips that do not qualify as tourism trips in the demand survey either due to frequency, purpose or duration of the trip (see **Figure 9**).

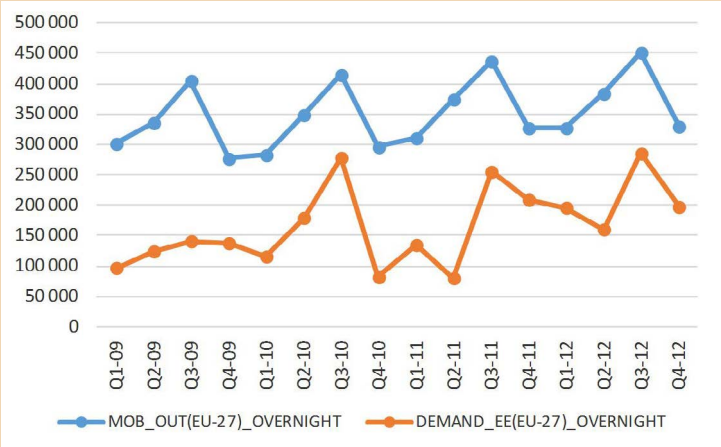

Figure 9. Quarterly outbound overnight trips (Estonians to EU27) in mobile positioning data when compared to the number of overnight trips to EU27 countries in Estonian demand statistics (*Source:* Positium, Statistics Estonia).

Further evidence for this is found by using the Finnish Border Interview Survey as reference data. Many Estonians commute to Finland making Finland part of their usual environment. The over-coverage in mobile positioning data is highest in the first and fourth quarters, both of which are outside the summer holiday season, which indicates that part of outbound trips in mobile positioning data are made by these commuters (see **Figure 10**). By definition commuters should be excluded from tourism statistics, so further filtering should be carried out for outbound mobile positioning data that is based upon visit frequency. This should be possible in case the data contains a non-changing subscriber ID.

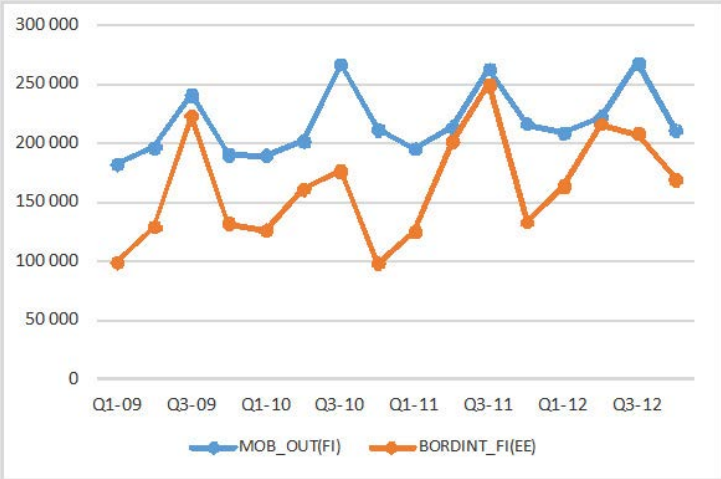

Figure 10. Quarterly outbound trips (Estonians to Finland) in mobile positioning data when compared to the number of Estonian visitors in the Finnish Border Interview Survey (*Source:* Positium, Statistics Finland).

One of the weaknesses of mobile positioning data is the potential misclassification of same-day and overnight trips (see **Figure 11**). At least two events are necessary at long enough time intervals (on different calendar dates) for a trip to be classified as an overnight trip. If only one activity is registered for the visitor during the entire trip, it will be considered a same-day trip in mobile positioning data although in reality the visitor may have spent one or more nights during their trip without using the phone. The analysis using Finnish demand statistics provides quantitative evidence of this misclassification. Mobile positioning data clearly overestimates the volume of Finnish same-day inbound trips to Estonia and underestimates the volume of overnight trips.

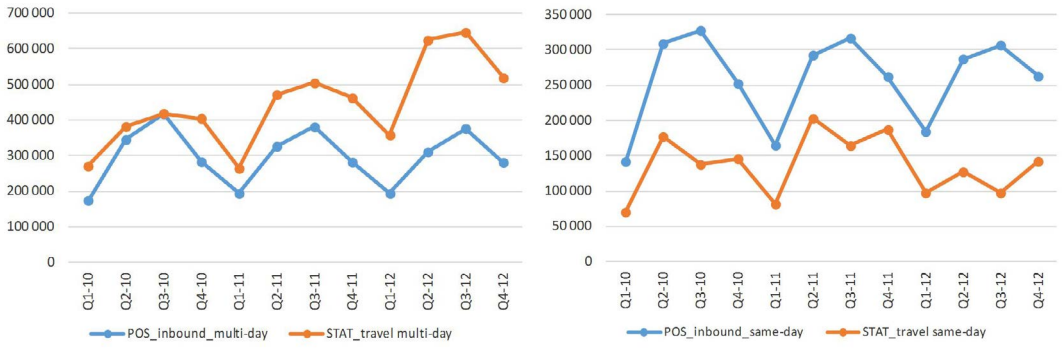

Figure 11. Quarterly inbound same-day and overnight trips (Finns to Estonia) in mobile positioning data when compared to the total number of same-day and overnight trips to Estonia in Finnish Demand Statistics (*Source:* Positium, Statistics Finland).

Domestic tourism differs from inbound and outbound tourism in the sense that the concept of more precise identification of usual environment within the country of reference is critical for domestic trips. Most of the trips shown in inbound and outbound mobile positioning data are part of tourism activities, with the exclusion of very short or very frequent trips. In domestic tourism the opposite is true: only a fraction of the total number of trips included in mobile positioning data are relevant for tourism activities due to the concept of usual environment. Also, the defining criteria for usual environment can be interpreted differently when compared to the official definition (Eurostat 2013). When comparing Estonian domestic mobile positioning data against accommodation statistics, the trends shown in overnight trips coincide with arrivals in accommodation statistics (see **Figure 12** and **Figure 13**). For demand statistics, when making a comparison against Estonian tourism demand statistics, the difference in domestic overnight trips is relatively small. It must be noted that due to the open interpretation of the criteria for identifying one's usual environment based on mobile data, the number of trips taken outside the usual environment can differ significantly.

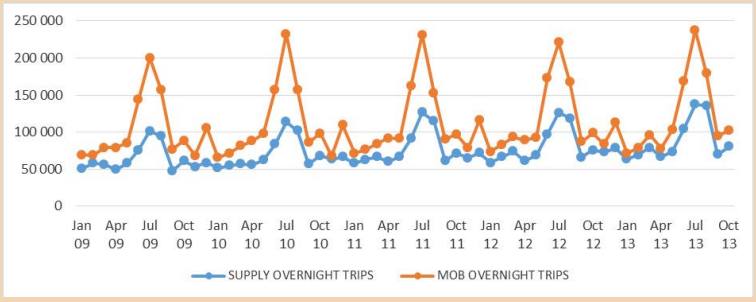

Figure 12. Monthly Estonian domestic corrected overnight trips in mobile positioning data when compared to accommodated domestic tourists in accommodation statistics (*Source:* Positium, Statistics Estonia).

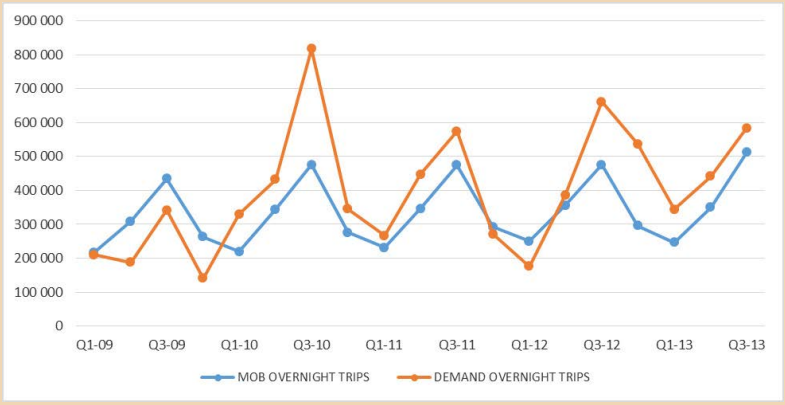

Figure 13. Quarterly Estonian domestic corrected overnight trips in mobile positioning data when compared to overnight trips made by Estonian residents in tourism demand statistics (*Source:* Positium, Statistics Estonia).

Regarding overnight trips, mobile positioning data provides a consistent estimate over time and the correlation to existing tourism statistics, both in terms of demand and supply, is very good for the most part. For example, a weaker correlation to supply statistics that involves Latvian visitors to Estonia may in fact indicate one of the strengths inherent in mobile positioning data as it captures a more complete tourism flow, as compared to supply statistics that only cover stays at paid accommodation. While mobile positioning data cannot provide the split into paid and non-paid accommodation (a problem only if the split as such is relevant for the users), it is nevertheless a more comprehensive data source for overnight trips.

A potential weakness is the presence of cross-border commuters in mobile positioning data. According to definitions used in tourism statistics, these commuters should be excluded from tourism statistics since they regularly commute to another country. Many Estonians commute to Finland regularly and might appear in the Estonian outbound mobile positioning data as seen in the analysis comparing mobile positioning data to data from the Finnish Border Interview. The exclusion of these commuters requires a non-changing subscriber ID in mobile positioning data based upon which the frequent commuters could be excluded from the inbound and outbound mobile positioning data (by the definition of the usual environment). As the raw data contains

this kind of user ID, such filtering of frequent visitors could be carried out at the level of raw data.

In terms of coverage and consistency, mobile positioning data tends to produce higher absolute figures when compared to the single reference statistics, although with some exceptions. Specifically, inbound overnight statistics based upon mobile positioning data tend to produce higher figures than the respective reference sources. As it can be assumed that traditional sources tend to underestimate real tourism flows, it can be argued that mobile positioning data provides a more realistic picture of complete tourism flows.

In addition, mobile positioning data in many cases show a variation over time ('consistency'), which is highly comparable to that of the respective reference statistics. Only in some cases is consistency lower. This is particularly true for domestic tourism activities.

Based upon the requirements laid out in Regulation 692/2011, it is clear that mobile positioning data alone cannot fulfil these requirements. Although mobile positioning data provide fairly good estimates for the number of trips, the nights spent and the destination, they do not produce any additional data about trips such as the purpose of the trip, the type of accommodation or the expenditure. For this information, additional data sources such as traditional surveys would still be needed.

Due to better timeliness when compared to official statistics, mobile positioning data would be particularly useful when it comes to providing quick indicators on the total number of tourism flows in order to complement the current supply statistics and their focus on paid accommodation alone. Mobile positioning data could also be used for producing more accurate regional tourism indicators. Mobile positioning data provides a source for calibrating the current surveys used for measuring tourism activities.

2.5. Opportunities and Benefits

Report 4. Opportunities and Benefits evaluates the opportunities and benefits that can arise from using mobile positioning data for tourism statistics. Data sources for this assessment include actual usage cases (as presented in *Report 1* of this study), quality, synergy and coherence assessments that are based upon real data (as outlined in more detail in *Reports 3a* and *3b*) and evaluations on data access, both with regard to technology and privacy (as described in more detail in *Report 2*).

The assessment of opportunities and benefits in this study touches five different perspectives: quality, cost, access to new indicators, synergies and transmission.

Within the **quality** perspective, it is outlined whether mobile positioning data is superior or inferior to more traditional data sources and processes in terms of completeness, timeliness, validity, accuracy, consistency and resolution. As a reference, tourism statistics as laid down in Regulation 692/2011 were chosen because this is the main framework for tourism statistics at the EU level today. It could be shown that mobile positioning data is highly consistent with reference statistics over time. At the same time mobile positioning data can be made available much more quickly than data from traditional sources, once the necessary processes are in place. As a result it can be shown that within the framework of official tourism statistics, mobile positioning data can be exploited as quick indicators and as calibration source. Of these two, quick indicators have the biggest potential to improve tourism statistics. On the other hand, it has to be stated that at present, mobile positioning data can hardly replace existing indicators within the framework of Regulation 692/2011.

Mobile positioning data can be used to potentially strengthen current tourism demand surveys through mixed-mode data collection. In such a scenario the number and duration of trips are based on mobile positioning data while tourism expenditure and ratios (purpose of trip, type of accommodation, means of transport, etc.) still rely on demand survey. The sample size of the demand survey could be decreased considerably since the survey does not need to support breakdown by destination, thereby reducing the cost and burden of data collection. Also more countries and even sub-regions could be included in the statistics since the sample size is not an issue in mobile positioning data.

Within the framework of the official regulation, the **cost** that would be incurred within an NSI when using mobile positioning data was assessed as opposed to the sources and processes exploited today. As a result it can be shown that using mobile positioning data would require between 168 and 264 man-days per year. When compared to the workload-induced traditional data sources and processes, mobile positioning data can be obtained and processed rather more efficiently. It has to be taken into account, however, that mobile positioning data is useful primarily as a calibration source or as a quick indicator, implying that at present, such data will be used in addition to other sources and processes.

Statistics as described by the Regulation 692/2011 are of course not the only tourism statistics that are being produced in EU Member States. The National Statistical Institutes (NSI) and other institutions produce tourism-related statistics that go beyond the framework of the regulation. Therefore, any possibility to gain **access to new indicators** through the use of mobile positioning data were assessed, firstly within official national tourism statistics (but not necessarily within the framework of the EU regulation), and secondly outside official tourism statistics (e.g. via destination marketing organisations). In official tourism statistics outside the EU regulation, mobile positioning data can deliver new information on those aspects, which are not required today mainly for methodological reasons. The two main fields of application are: 1. on the supply side, overnight tourism in accommodation establishments below the threshold and 2. on the demand side, overcoming the restriction of tourism for personal reasons. Mobile positioning data can reflect tourism in all sorts of establishments (including private accommodation) and for all purposes (including business). Spatial and timely resolution can, on average, be much more refined than required by the regulation when using mobile positioning data and therefore open up new possibilities when it comes to tourism statistics. When compared to surveys, mobile positioning data shows a number of advantages in terms of accuracy (smaller sampling error, no memory gaps), regardless of the selection bias.

Outside the field of official tourism statistics, mobile positioning data can be exploited in various contexts, such as detailed statistics in time and space, volume and structural breakdowns (nationality) for big events and other related applications, statistics describing accommodation not being covered in tourism statistics or segmentation data relying on subscriber master data.

Assessment of **synergies** also goes beyond the scope of the regulation on tourism statistics to show uses in other domains. It has been shown in pilot studies that mobile positioning data can be used in terms of official statistics specifically in the travel item of the Balance of Payments in transport and commuting statistics. Especially for transport statistics, mobile positioning data has the potential to provide new insights that are otherwise unavailable. For an NSI, the decision to exploit mobile positioning data within the regulatory framework of the specific country will be much easier when it becomes possible to share costs over different statistical domains.

Assessing the opportunities that might arise from the **transmission** of data can be viewed within the framework of the regulation (i.e. from NSIs to the European Commission), but little impact is to be expected in this area. A higher impact will probably arise from the transmission of data from MNOs to the NSI — but this only after the necessary automation processes have been carefully planned and thoroughly tested.

3. Implications for the Statistical Community and Users of Tourism Statistics

At the moment, the implications for tourism statistics at European level, as governed by Eurostat and covered by Regulation 692/2011, are rather limited. The main reason is that the heterogeneity of rules and regulations concerning access to mobile positioning data does not allow for useful application in all Member States. Within the consortium, countries with very liberal regulations on data protection (as in Estonia) and fairly strict regulations (as in Germany) have been covered. It could be shown that the technical possibilities offered by the use of mobile positioning data in tourism statistics and related fields obviously become merely practical in those countries in which access to the data is possible.

However, there are some opportunities tied to these implications even at the European level:

- The members of the European Statistical System (ESS), particularly the NSIs, can profit from the large number of experiences and usage cases contained and analysed in this study;
- Eurostat can follow the development that will certainly take place in the coming years in selected Member States and therefore be prepared to implement the system when the time has come;
- Eurostat, therefore, can function as a moderator for rolling out the use of mobile positioning data on the national and regional level, as has been the case for other projects (e.g. TSA, sustainability indicators).

At the national level, the possibilities available when it comes to tourism statistics, related statistics and other applications related to tourism activities are dependent upon two main factors:

- The accessibility to data: as shown above, at the moment there is a broad range of regulations where privacy and access to personal data are concerned. It is up to the NSI to verify their country's position on the continuum from 'very liberal' to 'very restricted';
- The specific situation for the sector that is intended to be monitored using mobile positioning data.

With regard to the second point, it could be shown that specific geographical situations (such as the ferry connection between Estonia and Finland over the Gulf of Finland) can facilitate the assessment and implementation of mobile positioning data based statistics. On the contrary, some existing geographical situations may render the use of such data particularly hard (e.g. a high share of migrants producing roaming data without being part of any tourism activities, as might be the case in Luxembourg).

An assessment of possible future opportunities and benefits would be incomplete without taking into account the trends in user behaviour and the technological advancements that are to be expected in the near future. One of these trends is the increasing usage of internet access through mobile phones (in addition to or even replacing speech and text messages), while a second trend is the possibility of tracking mobile devices not only through their position relative to network antennae, but rather through built-in satellite positioning systems. Both trends need to be monitored closely for future assessments of the usability of mobile data for statistics.

Future developments in the area of the usage of mobile positioning data are not easily foreseeable in detail, but the increasing levels of interest and the number of related projects shows that this data source is being increasingly used. The role of Eurostat, NSIs and tourism marketing organisations in this can be considerable, especially if there is a clear future vision. Although this is not the purpose of the current study, a simplified bold speculation could be proposed:

- Eurostat is leading the development of the methodology, concepts and definitions on tourism statistics indicators compiled from mobile positioning data. The approved methodology is applied in countries where mobile positioning data is used for producing statistics;
- The NSIs are the central collectors of the data from MNOs and the main dissemination point for the users of the data. The process is fully automated excluding manual quality assurance and extends from initial data collection to transmission to Eurostat and other dissemination channels (e.g. tourism observatory). The processes are almost simultaneous, meaning the results are transmitted to all channels after they are approved and released by responsible NSI;
- An alternative option, which is perhaps more complicated to implement but essentially more effective and cost-efficient, could be to have a supranational level organisation or body with explicit purpose, powers and legal framework for gathering, processing and disseminating mobile positioning-based statistical indicators on tourism from national or global MNOs;
- The data is used for tourism and other official domains where the value of mobile positioning data is beneficial;
- The scope of official statistics is expanded or the NSIs provide a wider scale of statistics for the users;
- The research for new opportunities in using the mobile data is carried out in cooperation with research institutes, private companies and MNOs. NSIs play a vital role in leading the research towards the needs of society in general and the inherent possibilities of these new data sources;
- MNOs are active participants in the ecosystem of statistics, and the legal framework allows them to benefit from additional revenues when it comes to commercialising the 'off-scope' data, such as in the spatial marketing sector;
- The ongoing goal for NSIs is to improve the quality, timeliness and variety of statistical products;
- The harmonisation of the output based on mobile positioning data is needed to guarantee comparability between countries;
- There is active participation in developing the education that supports the field (in social sciences and related disciplines such as statistics, computer science, geospatial analysis, and economics), the goal being to ensure the data scientists and statisticians of the future are being educated.

Such a vision certainly requires a reshaping of the role of the NSIs, the political will to enact such a change, and an understanding of how much added value this provides to society when compared to the costs that are involved.

Acknowledgments

The Consortium members would like to thank to everybody who helped with wisdom, knowledge and advice during the project. Our special thanks go to Estonian MNOs who made it possible to conduct the analysis based on their pilot data but also other MNOs who were willing to investigate the potential and participate in the discussion on the possibilities and obstacles of using mobile data. We also thank the respondents of the online survey and the interviewees who helped us greatly in mapping interest concerning the data and issues that are faced by the stakeholders. We are deeply hopeful that the study will be of great use for all interested parties and will help in understanding the different aspects of the use of mobile positioning data.

References

European Commission (2009) COM(2009)404 — Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF)

Eurostat (2012). 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. 2 May 2012 (http://ftp.infoeuropa.eurocid.pt/database/000048001-000049000/000048731_2.pdf)

Eurostat (2013). Methodological manual for tourism statistics, Version 2.1. Cat No: KS-GQ-13-007-EN-N (http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-GQ-13-007/EN/KS-GQ-13-007-EN.PDF)

KS-32-13-400-EN-C

Publications Office

ISBN: 978-92-79-39762-2
doi:2785/55051