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

Eurostat Contract No. 30501.2012.001-2012.452

Report 4. Opportunities and Benefits

15 April 2014

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

Eurostat contract 30501.2012.001-2012.452

Report 4. Opportunities and benefits

Final report

15 April 2014

Authors of this report (alphabetically):

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, task coordinator), **Ulf Sonntag** (NIT), **Margus Tiru** (Positium, project coordinator)

Table of Contents

1 Summary

In this report, we evaluate 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 report touches five different perspectives: quality, cost, access to new indicators, synergies, and transmission.

Within the quality perspective we outline 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 frame of reference for this area we chose tourism statistics as described by Regulation (EU) 692/2011 because this is the master framework for tourism statistics at the EU level today. As a result we can show that mobile positioning data can hardly replace existing indicators within the framework of Regulation (EU) 692/2011. However, it could be shown that mobile positioning data is highly consistent with reference statistics over time. This is particularly true when it comes to inbound tourism flows, while domestic tourism still has a major drawback in terms of determining the usual environment within mobile positioning data. Consistency over time in combination with a considerably better timeliness make mobile positioning data specifically useful for quick indicators on selected aspects of tourism (with a focus on inbound tourism activities).

Also within the framework of the official regulation we assessed the cost that would be incurred when using mobile positioning data as opposed to the sources and processes that are exploited today. As a result we could show that using mobile positioning data would require between 168 and 264 man days per year. When compared to the workload that is incurred by 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 such data will be used in addition to other sources and processes.

Statistics as described by the regulation 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, we assessed any possibilities that might be involved in gaining access to new indicators through the use of mobile positioning data, firstly within official (NSI-based) tourism statistics (but not necessarily within the framework of the regulation), and secondly outside official tourism statistics (e.g. through destination marketing organisations). As a result for this dimension we could show 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. Outside the field of official tourism statistics, mobile positioning data can be exploited in various contexts, such as for detailed statistics in time and space, volume and structural breakdowns (nationality) for big events and other related applications; statistics that describe accommodation not being covered at all in tourism statistics or segmentation data relying on subscriber master data.

Also beyond the scope of the regulation is the assessment of synergies. It has been shown in pilot studies that mobile positioning data can be used for other domains in terms of official statistics, specifically in the travel item of the Balance of Payments, in transport and commuting statistics. Specifically 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 Mobile Network Operators (MNOs) to the NSI - but this only after the necessary automation processes have been carefully planned and thoroughly tested.

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.

Altogether, we can conclude that mobile positioning data offers high levels of potential in terms of providing quick indicators for selected statistical information (such as the number of inbound tourists). Furthermore, mobile positioning data can yield additional information, specifically in terms of showing complete tourism flows (as opposed to only segments in today's supply and demand-side statistics), in finer spatial and timely granulation and even in the provision of longitudinal data.

All of these opportunities and potential benefits when it comes to using mobile positioning data are, however, restricted to the boundaries that are set by law (specifically when it comes to privacy), and the economic interests of the parties involved.

2 Introduction

This report concentrates on the potential opportunities and benefits of the usage of mobile positioning data for tourism statistics. Overall, in task 4 the project team did not collect and research new data, but rather integrated the results from previous work packages into a structured and coherent assessment of potential opportunities and benefits.

The report, of course, is not simply a rearrangement of the output of tasks 1-3. Its added value comes from the structural qualitative and quantitative assessment of the five key dimensions mentioned in the Terms of Reference (ToR, see also Table 1). However, when it comes to working with empirical data, this report is drawing on empirical results from other reports within the scope given there.

It becomes clear that the focus of this report is on opportunities and potential benefits, but not on possible financial, technological and legislative barriers when it comes to the use of mobile positioning data for tourism statistics. Therefore, this report will lead to an overall more positive impression when compared to an assessment report that would include both positive (opportunities or potential benefits) and negative (threats or risks) aspects of the use of mobile positioning data for tourism statistics. Potential threats or risks (as outlined in the ToR) are access to databases and various methodological issues and are discussed in the Reports 2, 3a, 3b and 5.

Table 1. Assessment dimension and description expected in Task 4 (source: Terms of Reference).

Assessment	Description (from ToR)	Based upon

Dimension		work from task
Quality	Potential gains in overall data quality, with special attention on the quality dimension, 'timeliness'	3a, 3b (coherence)
Cost	Potential reduction in burden/cost for, on the one hand, respondents and, on the other hand, administrations (e.g. statistical offices); this could include recommendations on how results that are obtained from mobile positioning data can be integrated into existing tourism statistics (to complement and/or replace data)	2 (for data access)
Access	Possible availability of or access to statistics/indicators which were not previously available (or which it is not possible to produce in a cost-efficient way)	3a
Synergies	Possible synergies with related domains (i.e. using the same sources as primary or auxiliary information); this should at least include a discussion on travel statistics (i.e. a travel item on the Balance of Payments)	3a
Transmission	The feasibility of automating the transmission of indicators to the Commission	2

These five dimensions are also mainly structuring the report. To assess these five dimensions, the report relies on the results of the previous tasks. All consortium partners worked on the other reports as well as on this report, so this report is a team authoring effort coordinated by NIT.

2.1 Aims of the Report

Primary aim of this report is to identify and assess potential opportunities and benefits of using mobile positioning data for different types of users and producers of tourism statistics and related fields.

The ToR outline for this task: 'The expected result for this task is a detailed report discussing all opportunities and/or potential benefits from using mobile positioning data for tourism statistics, when compared to the currently used production process, as well as the extension into related domains. The discussion can be from a theoretical/scientific point of view before moving to a practical/empirical point of view (on the basis of actual datasets and tests conducted for the feasibility study)'.[1]

2.2 Reference to Other Reports

This report follows Reports 1 to 3b. Consequently, the previous Reports 1 to 3b provide input data, information and expert opinions for this report, while Report 5 receives

[1] As outlined before, this report draws on the empirical results shown in previous reports and is therefore limited to the datasets that are available for analysis in previous reports.

input from this report (Table 2). More concrete references between the reports can be found in the descriptions of Chapter 3 and its sections.

Table 2. References to Report 4 found in other reports.

Report (from kick-off meeting presentation)	**Main reference to Report 4**
Report 1 − Stocktaking: an overview of literature and publications, an overview of applications and projects, an in-depth list of research projects and applications, an expert survey and in-depth interviews, charting the progress of accessing mobile positioning data, and an overview of legal, technical and methodological details.	Overall input for Report 4
Report 2 − Feasibility of access: privacy and regulation-related barriers, financial and business-related barriers, technological barriers; Report 2 results in a detailed report of the regulatory, financial, technical and other related topics that cover the aspects of data accessibility in EU countries	Provides input for assessment dimensions: - the cost of data access - transmission
Report 3a − Feasibility of use, methodological issues: identification of, description for, and possible solutions to a number of issues in a long chain of raw data processing so that it can be included in qualitative statistical results	Provides input for assessment dimensions: - quality - access - synergies
Report 3b − Feasibility of use, coherence: an evaluation of coherence between official statistical indicators on the national and regional levels and mobile positioning data-based estimations and results.	Provides input for assessment dimensions: - quality
Report 5: Consolidated report	Receives input from Report 4 (as well as all other tasks)

2.3 Scope of the Report

The main scope of the report is on assessing the use of mobile positioning data for tourism statistics as outlined in Regulation (EU) 692/2011. Here, the two most important dimensions ‘quality’ and ‘cost’ will be assessed (plus Transmission). Cost, together with access, will also be assessed for Official Tourism Statistics outside the regulation (e.g. replacing border surveys through mobile positioning data) and in other fields that belong to the tourism sector (e.g. the use of mobile positioning data by NTOs). The relationship between official statistics for tourism and other sectors concerning the use of mobile positioning data will be assessed in the synergies chapter (Figure 1).[2]

[2] As an additional reference point, the usage cases discussed in the report on Task 1 in this project have been sorted into the scoping matrix fields - see Annex 1.

Scoping matrix

	Official Statistics (EU level)	**Official Statistics** (national level)	**Other fields of application**
Tourism sector	**Scope T1**: Statistical sectors as covered by Regulation (EU) 692/2011 ⇒ *Quality (2.1)* ⇒ *Cost (2.2)* ⇒ *Transmission (2.5)*	**Scope T2**: Additional tourism statistics produced by National Statistical Institutes ⇒ *Cost (2.2)* ⇒ *Access (2.3.1)*	**Scope T3**: Additional fields of application in tourism outside official statistics ⇒ *Cost (2.2)* ⇒ *Access (2.3.2)*
Other sectors	**Scope O1**: Relevant sectors of official statistics outside tourism ⇒ *Synergies (2.4*)		**Scope O2**: Other sectors and other fields of application *(descriptive)*

Figure 1. Scoping matrix for Report 4.

As outlined above, the five assessment dimensions (see Table 1) mainly structure the report. The assessments for the quality, cost and transmission dimension are applied to the relevant tourism statistics sectors as covered by Regulation (EU) 692/2011 (Table 3).

Table 3. Tourism statistics, sectors to be covered.

Sector	As described in	Remarks, further explanations
a) Accommodation capacity	Regulation (EU) 692/2011, Annex 1, Section 1	Probably the least relevant, but covered nonetheless for reasons of completeness
b) Occupancy of tourist accommodation establishments	Regulation (EU) 692/2011, Annex 1, Section 2	Arrivals and nights spent, be region and nation, by residents vs. non-residents
c) Internal tourism in non-rented accommodation	Regulation (EU) 692/2011, Annex 1, Section 4	Although it is very clear that mobile positioning data can hardly distinguish between different types of accommodation, this sector should be covered for reasons of completeness.
d) National Tourism (overnight)	Regulation (EU) 692/2011, Annex 2, Sections 1 and 2	by duration and country/locality of residence, month, country/locality of destination, type of destination, etc.
e) National Tourism (same day visits)	Regulation (EU) 692/2011, Annex 2, Section 3	outbound vs. domestic
f) Additional sectors	-	*reserved space for potential new indicators related to tourism statistics arising from the methodology which cannot be sorted into any of the existing sectors*

3 Results: Opportunities and Benefits

This chapter presents the findings on opportunities and potential benefits in using mobile positioning data, when compared to the traditional data sources and production processes being employed today. This chapter is subdivided into the five assessment dimensions:

1. The quality dimension, which assesses relevant quality indicators along the sectors that are outlined in Regulation (EU) 692/2011;
2. The costs dimension, which assesses possible cost effects along the Generic Statistical Business Process Model (GSBPM);
3. The access dimension, which reveals additional possibilities when it comes to using mobile positioning data outside tourism statistics as covered in Regulation (EU) 692/2011;
4. The synergies dimension, which reveals additional possibilities that may be available in using mobile positioning data outside tourism activities;
5. The transmission dimension, which reveals additional possibilities when it comes to using the sectors that are outlined in Regulation (EU) 692/2011.

The numbering above is the same as the numbering of the sections.

3.1 Quality Dimension

This chapter shows, if and how different aspects of data quality can be improved by using mobile positioning data. The point of reference is either absolute (i.e. the assessment reflects the use of mobile positioning data alone) or relative to today's situation (i.e. the assessment is compared to today's situation). Aspects to be assessed are shown in Table 4.

Although there is no Total Quality Index for (tourism) statistics today, there is a *Rolling Review on Tourism Statistics* (Eurostat 2013b) providing valuable framework information on user needs, possible shortcomings of today's statistics and perspectives of European tourism statistics for the future. The Rolling Review states, among others, as particular issues in today's European tourism statistics:

- Different definitions in the Balance of Payments travel concept and the tourism statistics visitor concept and coherence of business statistics and accommodation statistics (p. 5);
- User's need for new information regarding economics (TSA), segmentation, operators and sustainability (p. 7);

- A user rating showing never more than 63% of '(very) good' in any quality dimension and a specific user's interest in timeliness ('provisional data') and coherence in terms of data collection (p. 8);
- Main recommendations going into the direction of more cost-effectiveness, reducing burdens and new statistics (p. 12) and harmonisation of registers, Europe-wide surveys, economic contribution (TSA) and improving timeliness (p. 13).

The assessment in this report draws partly on this framework information and furthermore is based upon our expert assessment (partly covered in previous reports) and the expert interviews and survey.

Table 4. Quality aspects to be assessed.

Quality aspect	Description
Completeness	Does mobile positioning data cover the indicators required for the relevant statistical sectors completely or simply more completely than today's statistical indicators? Are the statistical sectors in the official regulation covered by all or only part of the indicators?
Timeliness	Can mobile positioning data be obtained, processed and deployed faster than today's statistical indicators?
Validity	To what extent does the available mobile positioning data serve as valid indicators for the real world facts that are to be covered in the relevant statistics sectors? (Does the data represent the facts that it is intended to represent?) *This aspect has been covered in detail in Report 3a, Section 3.1*
Accuracy	Does mobile positioning data cover the indicators required for the relevant statistical sectors accurately or simply more accurately than today's statistical indicators? *This aspect has been covered in detail in Report 3a, Section 3.2*
Consistency (integrity, coherence and comparability)	To what extent is mobile positioning data consistent over time (continuity) and spatial units (comparability)?; *this aspect has been covered in detail in Report 3a, Section 3.1* To what extent is mobile positioning data coherent with other sources? *This aspect has been covered in detail in Report 3b*
Resolution	To what extent can mobile positioning data be delivered in a higher resolution/granulation when compared to today's data (resolution referring to the dimensions of space and time)? *This aspect has been covered in detail in Report 2, Section 3.1*

When compared to the output quality dimensions in use by the ESS (2012), the list varies only in some detail:

- Relevance does not need to be assessed here because we are working in the framework of Regulation (EU) 692/2011. We assume that the regulation only contains relevant indicators. This does not imply, however, that all possible relevant indicators are covered in the official regulation.[3] Therefore, additional indicators are being discussed in this report in the subsequent chapters.

[3] 'Relevance is the degree to which statistical outputs meet current and potential user needs. It depends on whether all of the statistics that are needed are produced and the extent to which

- Accuracy and Reliability (covered in the list under accuracy);
- Timeliness and Punctuality (covered in the list under timeliness);
- Coherence and comparability (covered in the list under consistency);
- Accessibility and clarity are not covered in the list, because these dimensions are mostly irrelevant for a feasibility study dealing primarily with data sources.

When compared to the quality indicators put forward in Regulation (EU) 1051/2011, again the list varies only in minor detail:

- Relevance, including completeness does not need to be assessed here because we are working in the framework of Regulation (EU) 692/2011. We assume that the official regulation only contains relevant indicators. However, it can be doubted that the official regulation is completely covering the information needs of relevant stakeholders and it might be the case that but there are 'under-covered' parts that are often due to limited feasibility of collection with the current methods. These aspects are covered in the chapter entitled 'Access to new indicators', below;
- Accuracy, including coverage-error (covered in the list under accuracy);
- Timeliness (covered in the list under timeliness);
- Punctuality (covered in the list under timeliness);
- Accessibility and clarity are not covered in the list, because these dimensions are mostly irrelevant for a feasibility study dealing primarily with data sources;
- Comparability (covered in the list under consistency);
- Coherence (covered in the list under consistency);
- Cost and burden are not in this list because they will be assessed in a separate chapter.

We can conclude that the list of quality indicators is in line with relevant European sources on the same topic and there are reasons for the few deviations.

3.1.1 Completeness

3.1.1.1 Coverage in Population

Before discussing the question of indicator-related completeness it is useful to have a look at mobile phone usage in the population in order to assess the theoretically possible completeness. At a first glance, the data seem to be very promising: for 2009, Eurostat

concepts used (definitions, classifications, etc.) reflect user needs.' (ESS 2009, p. 7).

reported: 'The average number of mobile phone subscriptions per hundred inhabitants stood at 125 in the EU27 in 2009. It surpassed parity (a hundred) in 24 of the EU Member States, where there were more subscriptions than inhabitants; the three countries in which rates were below a hundred subscriptions per hundred inhabitants were Austria, France and Latvia'.[4] However, this data does not do anything at all to reveal the share of the population that uses mobile phones (because it is obvious that one person can have more than one subscription, while other persons do not use mobile phones at all). Eurostat does not hold any statistics on this topic, and furthermore the telecommunication statistics published on the Eurostat database (tin00059 and tin00060) end in 2009. More up-to-date indicators are available from the World Bank (cf. Report 2, Section 4.6.1).

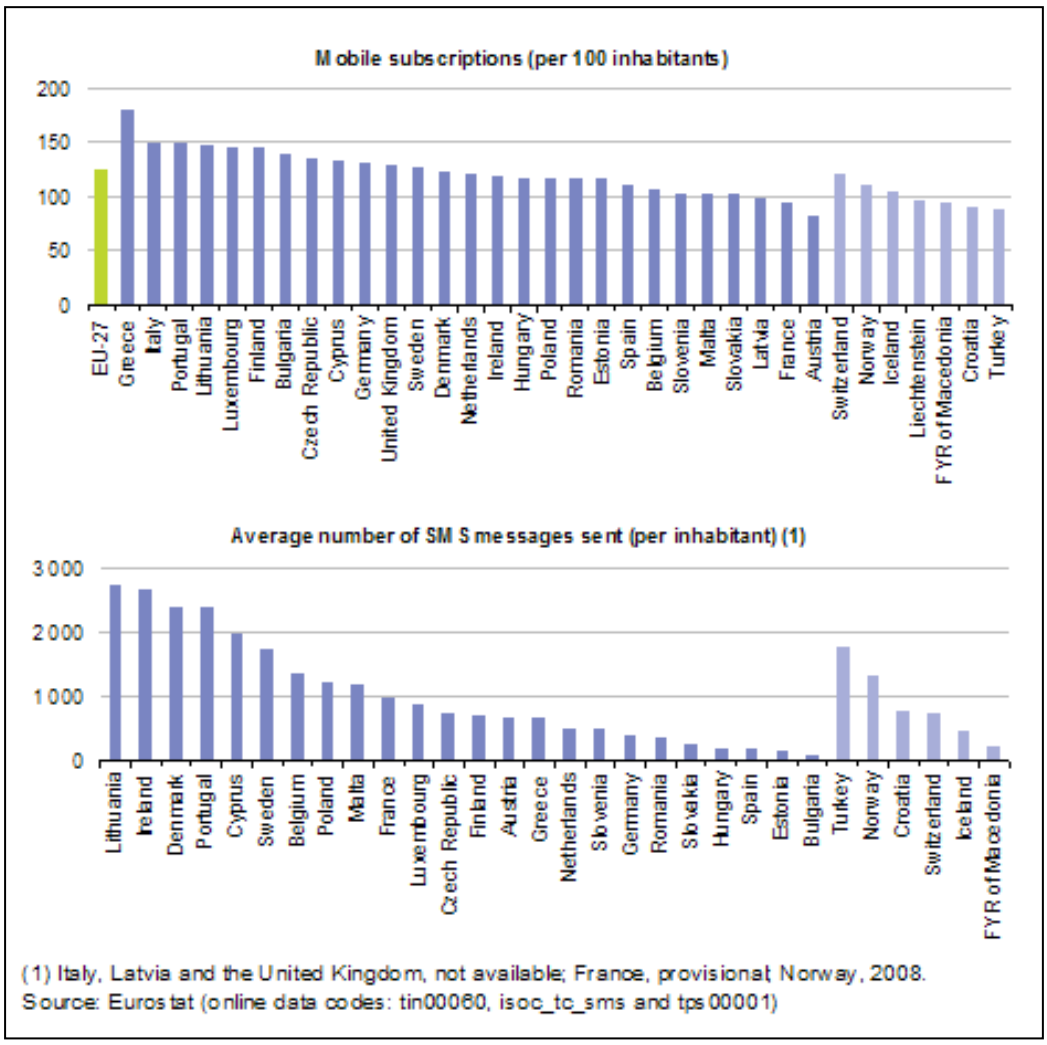

Figure 2. Mobile subscriptions per hundred inhabitants in the EU27, 2009 (source: Eurostat).

[4] Source: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Telecommunication_statistics

For example, for Germany a recent report (E-plus 2013) shows that in 2013 there were 113.2 million mobile phone subscriptions. A total of 2.3 million of these are for machine-to-machine-communication. The rest of the 110.9 million subscriptions gives us an average of 138 subscriptions per hundred inhabitants (with a total population of 80.5 million). However, 'only' 84% of the population have access to mobile devices. As a consequence, an average of close to or even more than one subscription per capita does not automatically imply coverage of 100% of the population.

It can be assumed that there is a positive correlation between mobile phone usage and participation in tourism activities. A person needs to have (among other factors) the resources to travel and people not travelling tend to adhere to specific socio-economic groups (e.g. lower education groups, see Table 5 for a German example). These will be the same groups who tend not to use mobile phones. However, precise data (from a single source and Europe-wide) is lacking in order to be able to test this hypothesis.

Table 5: Mobile phone usage and holiday travel in socio-economic subgroups, Germany.

Segment	Share of population using mobile phones in the last 3 months (2012) (%)	Share of population (14+) making at least one holiday travel in the last 12 months (2012) (%)
16-24 years	96	79
25-44 years	95	80
45-64 years	89	80
65 years and older	72	63
Level of education: low	76	65
Level of education: middle	86	82
Level of education: high	90	89
Employment status: employed	93	84
Employment status: unemployed	82	44
Employment status: pupil/student	97	84
Employment status: retired	69	64

Sources: Statistisches Bundesamt: *Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien.* Wiesbaden, 2012 (left-hand column), Forschungsgemeincshaft Urlaubs und Reisen: Reiseanalyse 2013 (right-hand column).

In conclusion, we can determine that probably a large proportion of the population in the European countries has access to mobile phones and is therefore registered by an MNO. This proportion, however, is most probably considerably below 100%, with actual and up-to-date figures being not known and differing from country to country.

As opposed to the discussion on validity, which covered the possible definitional issues in comparing mobile positioning data to traditional data sources, this chapter will focus on the actual completeness of indicators addressed in the official regulation.

We will therefore evaluate whether the statistical sectors and the indicators described within the sectors can be (better) covered by mobile positioning data when compared to traditional data sources.

3.1.1.2 Internal Tourism: Capacity of Tourist Accommodations (Annex 1, Section 1)

Obviously, because mobile positioning data always refers to a user and not to an establishment, little additional information is to be expected in this sector.

It can be technically possible, however, to identify spatial clusters of non-resident's mobile devices in an area where there is no statistically known accommodation establishment (related to the 'anchor-point' or 'meaningful places' technique described in Report 3a with regard to the identification of the 'usual environment'). This will surely not deliver any capacity information, but would allow statistical offices to manually look for any establishments not reporting today although they would have to according to statistical regulations. However, there are no automatic procedures in place and no pilot projects have been carried out in this regard.

3.1.1.3 Internal Tourism: Occupancy of Tourist Accommodations (Annex 1, Section 2)

The number of nights spent and the number of arrivals for residents and non-residents can theoretically be derived from mobile positioning data. The regional granulation can be finer than the required NUTS 2 level. The 'types of locality' can easily be differentiated (but not the types of accommodation - see below).[5]

In addition to the requirements of the official regulation, demand figures can obviously be derived from mobile positioning data regardless of a capacity threshold. When it comes to mobile positioning data, because there is no threshold that is related to the size of an establishment, the data includes all tourism flows. This will, however, include rented and non-rented accommodation without the possibility of being able to distinguish between these two groups.

[5] The 'types of locality' mentioned are: a) a densely populated area, an intermediate area, and a thinly populated area; b) coastal and non-coastal.

However, a differentiation between different sizes or types of accommodation (NACE 55.1, 55.2 and 55.3) is not possible. The data can be extracted monthly or for any other timeframe.

Occupancy rates cannot be derived from mobile positioning data because they need capacity information which cannot be drawn from mobile positioning data (see paragraph above).

3.1.1.4 Internal tourism: Non-rented Accommodation (Annex 1, Section 4)

Mobile positioning data does not allow for any differentiation between different types of accommodation (see paragraphs above). Therefore, non-rented accommodation cannot be identified using this data source.

However, the usage of mobile positioning data might provide fresh options. If User-IDs can be tracked for a longer period of time, it might be possible to identify regular mobility patterns which indicate the usage of non-rented accommodations (specifically second homes and VFRs - visiting friends and relatives). This possibility is, in practical terms, not available when using more traditional data sources.

Furthermore, it is possible to estimate that area of tourism flow that stays in non-rented accommodation, if used in connection with tourism demand and supply statistics. Report 3b has shown the possible over-coverage and under-coverage of mobile positioning data when compared to other sources in terms of tourism. It has been shown that mobile positioning data tends to overrate the level of tourism if it is compared against supply-side statistics. If it could be possible, however, to estimate the level of under-coverage in supply-side statistics when compared to complete tourism flows, then this factor can be applied to mobile positioning data and consequently provide the possibility of tracking changes in tourism flows over time. If this possibility is not provided (because reliable third party data is lacking), then it can be argued that mobile positioning data gives a more realistic view on actual tourism flows when compared to most single reference statistics (see also Section 3.1.5 on consistency).

3.1.1.5 National tourism: Participation in Tourism for Personal Purposes (Annex 2, Section 1)

Assuming that a Subscriber-ID can be tracked for a longer period of time (e.g. a year), it is technically possible to derive information on each and every trip which a person (using a mobile phone) takes during the given period (e.g. a year). The dataset from which this information is to be derived needs to contain information on at least the following points:

- Subscriber identity code;
- Usual environment (the calculation is based upon the spatio-temporal patterns that are taken from longitudinal data);
- Trip destination;
- Duration of the trip.

In addition to the subscriber identity code, further information on age, gender, educational level (optional), employment situation (optional) and household income (optional) would be necessary in order to be able to cut off at the age of 14 and fulfil the 'aged fifteen or over' requirement and additional breakdowns in the regulation (however, it is assumed that the age threshold is in the regulation mostly because of the limitations of interview-based surveys and does not necessarily have another factual background).

A possible drawback in using mobile positioning data in this field is the limitation on 'personal purposes'. Obviously, this aspect cannot be covered by using mobile positioning data because it is not possible to decide on the purpose of a trip from the data alone (except for some very basic assumptions, e.g. weekend trips - see Report 3b Section 4.4.1 - but certainly not for all trips). Even though the Methodological Manual for Tourism Statistics (Eurostat 2013b, Paragraph 3.2.1) states that 'trips made for professional reasons are excluded for practical reasons (they are more difficult to collect and to recall)' (which would make it an advantage to have the complete view through using mobile positioning data), there is also a relevance-related reason for the limitation to 'personal reasons', because participation 'touches the social component of tourism statistics, as such the inclusion of trips for professional reasons when studying the population's participation rate is of secondary importance or even not desirable'.

Consequently, although it is possible to make some assumptions based upon the subscriber's time-behaviour (for instance, if domestic tourists travel in the middle of the week then this is more probably a business-related trip while any weekend trips are personal ones), and in many cases it is also possible to determine the longer holiday periods (trips taken in the summer), and to connect trips that are made outside the usual environment as being personal holiday trips, a clear distinction between travel purposes cannot be provided by using mobile positioning data.

While the purpose of trip cannot be derived from mobile positioning data, it is not common for anyone to take trips for business purposes alone during the entire year. According to Finnish Demand Statistics, only 1% of respondents who took at least one

business trip during 2012 did not participate in tourism activities for personal purposes during the same year. In contrast, of those respondents who didn't take any business trips in 2012 more than 10% also didn't participate in tourism activities for personal purposes. Participation in tourism activities for personal purposes is therefore significantly higher for persons who take trips for business purposes.

According to Eurostat's 2012 tourism demand survey, the share of business trips in all overnight trips was 14% as an average for 22 reported countries and varied between 19% for Germany and 2% for Romania. It shows that if business-related trips are not discounted from mobile positioning data it leads to an overestimation of personal trips and corrections are recommended in order to avoid such an overestimation. However, corrections that are based upon the time and date of the trip may not be sufficient because the predictive power of such a model is unknown. The date of the start and end of a trip is not collected in the demand survey and therefore the hypothesis regarding the relationship between the type of the trip and the time of the trip cannot easily be tested.

3.1.1.6 National Tourism: Tourism Trips and Visitors Making the Trips (Annex 2, Section 2)

Under the same assumption, as before, some of the required data can be obtained and some of it cannot (Table 6).

Table 6. Information on tourism trips and visitors to be obtained from mobile positioning data.

Can in theory be obtained from mobile positioning data	Cannot be obtained from mobile positioning data
• Month of departure • Duration of the trip • Number of nights spent on the domestic territory • Main country of destination • (Type of destination) • Profile of the visitor (country of residence)	• Main purpose of the trip • Main means of transport • Main means of accommodation • Booking of the trip • Expenditures • Profile of the visitor (gender, age, educational level, household income, employment status) - *if not to be supplied by the MNO's master data*

3.1.1.7 National Tourism: Same Day Visits (Annex 2, Section 3)

Under the same assumptions, as before, it is possible to derive part of the necessary basic information from mobile positioning data, but a substantial amount of information that is required by the official regulation cannot be obtained by using mobile positioning data exclusively (Table 7).

Table 7. Information on same day visits to be obtained from mobile positioning data.

Can in theory be obtained from mobile positioning data	Cannot be obtained from mobile positioning data
• Number of outbound same-day visits • Country of destination (outbound) • (Number of domestic same-day visits)	• Purpose (personal vs. professional) • Expenditures • Profile of the visitor (gender, age, educational level, household income, employment status) - *if not to be supplied by the MNO's master data*

3.1.1.8 Overview and Conclusion

In conclusion, we can determine that from the fields that are covered by Regulation (EU) 692/2011 none can be completely covered by using mobile positioning data alone. Two aspects on national tourism, however, can at least be covered partly (Table 8).

Table 8. Information on tourism trips and visitors to be obtained from mobile positioning data.

Can in theory be completely covered by using mobile positioning data	Can in theory be partly covered by using mobile positioning data	Cannot be covered at all by using mobile positioning data
-	• National tourism: tourism trips and visitors making the trips (Annex 2, Section 2) • National tourism: same day visits (Annex 2, Section 3)	• Internal tourism: the capacity of tourist accommodation (Annex 1, Section 1) • Internal tourism: the occupancy rates for tourist accommodation (Annex 1, Section 2) • Internal tourism: non-rented accommodation (Annex 1, Section 4) • National tourism: participation in tourism for personal purposes (Annex 2, Section 1)

It has to be taken into account, however, that the requirements of the regulation follow to a certain extent the methodological toolsets available - in other words, what cannot be measured, will not be required by the regulation. Therefore, a certain natural disadvantage occurs for any data source other than those that are in place when the official regulation has been implemented. Therefore, although this chapter argues mainly within the framework of the current requirements of the official regulation, further opportunities that may arise from the usage of mobile positioning data will be discussed in subsequent chapters of this report.

3.1.2 Timeliness

Timeliness describes the time lag between the event described and the date of publication of statistical information.

The Rolling Review on Tourism Statistics (Eurostat 2013b) states timeliness as one issue from the user perspective in tourism statistics today ('For timeliness, the main requirement was for earlier provisional data, with an automatic notification of later updates', p. 8). Furthermore, the European Statistical Programme 2013-2017 (Regulation (EU) 99/2013) addresses timeliness in Lit. 16 of the preamble: 'Sixth, due to the appearance of new actors on the information market, including those providing information in nearly real time, the priority for the ESS in future is high quality, including timeliness'.[6]

Using mobile positioning data, real-time processing can theoretically be obtained if automatic procedures were to be installed. These procedures would have to allow for mostly automatic data collection, processing and deployment. Therefore, using mobile positioning data could provide statistics faster than is the case today.

However, there are three main drawbacks:

- Implementing automatic data processing schemes would mean having the technical and organisational means to do so. It would require a massive IT infrastructure to handle very large amounts of data (see Report 2, Chapter 4) and it would require online access to the MNO's infrastructure, which would probably mean overcoming financial, organisational and privacy barriers (see Report 2, Chapter 5);
- The second aspect is closely related to the first and refers to the cost of such an implementation. Chapter 3.2 of this report analyses the cost of using mobile positioning data in more detail. It is clear, however, that there is an exponential time-cost-curve: there is not much difference in terms of cost between retrieving data once a week or once a month, but there is a relatively large difference between retrieving data at two-hour intervals and 24-hour intervals (see Section 3.3.1.4 of this report for a more detailed elaboration on quick indicators);
- Thirdly, there is a methodological aspect to be taken into account. It refers to the fact that every report that covers mobility (tourism included) needs a minimum time lag in order that it can be sufficiently exact, simply because the 'act of mobility' needs to be finished to be reported correctly. This is true for every data source, such as in terms of surveys (people who are still in the process of conducting their trip cannot usually be surveyed). Using mobile positioning data, the problem is practically the same: in order

[6] The proposal for the programme stressed timeliness even more strongly: 'Thirdly, due to the appearance of new performers on the information market, including those providing information almost in real time, the priority for the ESS in the future is quality and, in the case of short-term statistics, timeliness in particular' (COM2011 (928), 21 December 2011, p. 8).

to identify a 'movement' for a mobile device as being a tourism-related visit (i.e. one that takes place outside the usual environment), the data concerning that certain period of time has to be collected and processed in order to distinguish the everyday 'movement' from general tourism activities and to specify the type of tourism activity involved (i.e. domestic or outgoing, day trip or overnight stay, etc.). For example, in every data update there are a proportion of trips that are not finished within the data period. If very short retrieving cycles are used to produce indicators, they would incur a need for *ex post* data correction.

As a consequence, timeliness can be improved by using mobile positioning data, although there are severe barriers to practically implement the technical and organisational processes needed mostly in terms of the cost of implementation and maintenance.

However, using mobile positioning data can contribute to the establishment of a 'quick indicator' for certain tourism statistics, e.g. the outbound travel flow or the inbound travel flow. Such data could be obtained in very short time intervals, and shifting the processing into the MNO's system would even allow figures to be drawn up on a daily or weekly basis (see Section 3.3.1.4 of this report for a more detailed discussion of quick indicators).

3.1.3 Validity

Statistical validity, in general, means the ability of an indicator or measurement to describe what it is intended to describe ('How well do the applicable concepts coincide with the intended concepts?'). Or in other words and more concrete for this study: 'To what extent does the available mobile positioning data serve as valid indicators for real world facts to be covered in the relevant statistics sectors?' 'Real world' facts in the framework of this chapter are basically those described in Regulation (EU) 692/2011. However, as pointed out above, mobile positioning data can provide access to new indicators covering different aspects of the 'real world' (please see subsequent chapters of this report, specifically Section 3.3).

The Report 3a shows a basic coherence of definitions between those outlined in the official regulation and those being applicable when using mobile positioning data (see Section 3.1 of Report 3a). The report investigates eleven concepts and definitions and shows that theoretically concepts are mostly comparable. The report states: 'Looking at the definitions, one can say that many of the listed differences are minor and their effect on the output is likely to be negligible (such as that concerning the duration of stays). Those concepts that

may have validity problems include the main destination (outbound), the usual residence and the usual environment (domestic), but the impact is difficult to assess'.

The main issue in assessing the validity of mobile positioning data seems to be the definition of 'usual environment', which plays a role in all forms of tourism activities, but has the biggest effect in measuring domestic tourism flows. Apart from that, differences between definition and measurable indicators seem to be minor.

3.1.4 Accuracy

To assess the accuracy dimension, three aspects can be taken into consideration:

- Coverage;
- Sampling;
- Measurement;
- Processing.

All of these aspects have been covered in detail in the Report 3a (Section 3.2). Therefore, we can restrict this section to the main findings contained in Report 3a.

3.1.4.1 Coverage

Report 3a states 42 possible source of overestimation and underestimation when it comes to indicators that use mobile positioning data, and certainly not all of those issues can be solved. As a consequence, the report states that 'many problems are inherent in the mobile positioning data and therefore cannot be avoided. Furthermore, their total effect, i.e. the total size of the coverage bias, needs to be evaluated or bias-corrected estimates need to be computed'.

This, at a first glance, makes mobile positioning data usable only after a careful evaluation and calibration using other sources has taken place. For example, as Report 3b has shown, overnight trips are better covered than same-day trips or short trips across the border.

In reverse, obviously, it makes clear that mobile positioning data can also be used to calibrate existing sources, because in the case of relevant differences between two sources it is not clear from the beginning which source produces which bias. In other words: in order to use mobile positioning data for the purpose of calibration, one needs to know its accuracy, and without it calibration is not recommended.

However, if the coverage bias is assumed to be stable over time, then mobile positioning data can still be used, - if not for level estimates then at least for change or trend estimates.

3.1.4.2 Sampling

A sampling bias in using mobile positioning data can only occur if samples are used. However, due to the possibility that large random samples can be drawn up, a sampling error can easily be computed and will become smaller the larger the sample is.

Therefore, when compared to surveys with (usually) smaller sample sizes and the problem of dropout and missing answers, mobile positioning data have a clear advantage.

3.1.4.3 Measurement

Report 3a makes clear that measurement issues and errors in the datasets can occur. However, there is no indication that these errors occur more frequently or have a higher impact when compared to traditional data sources. In surveys, measurement issues are frequent (coming from errors on the side of interviewees like memory gaps, interviewers or data typing and processing), and also in accommodation statistics is it to be expected that errors occur (e.g. intentional or non-intentional errors produced by the accommodation establishments).

As a result, measurement errors can occur, but are probably not more problematic as today.

3.1.4.4 Processing

One challenge in processing mobile positioning data is the massive amount of data to be handled. This, however, is not a systematic problem, but rather a question of hardware scaling and costs incurred (see the chapter on cost in this report).

For statistics on domestic tourism activities, determining the ‘usual environment’ is a challenge in the processing phase. Report 3a states the anchor-point-method to determine the usual environment (Section 2.3.1.1 and Section 3.2.4) with up to 95% correct identifications (on second administrative unit level, LAU2).

3.1.4.5 Conclusion

In conclusion, there are a number of issues with regard to the accuracy of indicators. These issues are less relevant in international tourism activities (inbound and outbound), but are relevant in domestic tourism activities because of the necessary definition of ‘usual

environment' from mobile positioning data datasets. These issues make it hardly possible to rely on mobile positioning data as the only source of information in tourism statistics.

3.1.5 Consistency (Integrity, Coherence and Comparability)

3.1.5.1 Consistency Over Time and Space

Report 3a extensively covers comparability over time and space in Section 3.3. It becomes clear that - when compared to traditional means of data collection - mobile positioning data can have some advantages. These advantages are specifically to be found in the fact that only a small number of data sources (i.e. the MNOs) are necessary to cover a multitude of cases. In traditional forms of data collection, several thousand sources (respondents, either accommodation establishments or consumers) need to be addressed. This is true for accommodation statistics (all accommodation establishments offering ten beds or more need to be contacted) and also for surveys (several thousand interviewees need to be addressed). In both cases, sampling bias can occur, with non-reporting being the most important. Using mobile positioning data usually means receiving data from up to five MNOs in order to cover a whole country.

Furthermore, as the basic technology for establishing mobile communication network is more or less comparable throughout the relevant countries and several MNO are having businesses in more than one country,[7] it can be conveyed that the necessary indicators can be delivered in more or less the same quality and format in the relevant countries. Report 2 states that 'from the point of view of data extraction for tourism statistics, a standard process would be to access data that is definitely stored by all MNOs according to standards' (Section 4.4)

On the other hand it must be conceded that relying on only very few sources might reduce sampling issues, but makes the system more vulnerable to the drop out of single sources. When analysing sensitivity, it becomes clear that one MNO that is no longer able or willing to deliver the required data means a lot more in terms of the problem of data consistency when compared to one interviewee or accommodation establishment dropping out of the traditional scheme.

[7] For example, the British Vodafone Group is active in more than twenty countries, Sweden's TeliaSonera in eighteen countries, Germany's Telekom in more than fifteen, and even smaller MNOs such as the Dutch KPN are active in five countries (considering only their own operations and substantial shareholdings).

3.1.5.2 Coherence

Report 3b has assessed the coherence between indicators that are based upon mobile positioning data and other data sources. Table 9 and Table 10 show the results with 'Average CC2AB' as an indicator for coverage (positive: mobile positioning data tends to be higher and shows over-coverage; negative: mobile positioning data tends to be lower and shows under-coverage when compared to the reference statistic) and 'r' as an indicator for consistency ('r' being in the range of +1.00 for perfect positive correlation to zero for no correlation and -1.00 for a perfect negative correlation).

As a result, we see in terms of coverage and consistency:

- Mobile positioning data tends to produce higher absolute figures (over-coverage) when compared to the single reference statistics, although with some exceptions (negative values in the first column of the table indicating under-coverage, shown in red). Specifically, inbound overnight statistics that are based upon mobile positioning data tend to produce lower figures than the respective reference source. 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;[8]
- Mobile positioning data in many cases show a variation over time ('consistency'[9]) which is highly comparable to that of the respective reference statistic. Only in some cases is the consistency lower (those below 0.90 appear in red in the table). This is particularly true for domestic tourism activities.

Table 9. Selected coherence indicators for inbound tourism activities (from Report 3b).

Segment covered	Coverage (Avg. CC2AB)	Consistency (r)	Reference statistics	Report 3b
FI-EE Inbound and Outbound (raw data)	-	+ 0.96	Ship transport	Table 6
FI-EE Inbound and Outbound (corrected)	- 0.16	+ 0.96	Ship transport	Table 7
SE-EE Inbound vs. Outbound (corrected)	- 0.01	+ 0.77	Ship/Air transport	Table 9

[8] In a recent study in the northernmost area of Germany, NIT found systematic under-coverage of supply-side statistics to vary between 5% and 40% in terms of capacity, leading to under-coverage in terms of demand as well. These figures only reflect under-coverage within the scope of the regulation (in this case establishments with ten or more beds), not taking into account the large remainder of establishments which fall outside the scope of the regulation.

[9] In this case, the consistency indicator 'r' reflects deviations in the data points in a one time series from the mean of this time series when compared between two time series. If both series data for a given month show simultaneous deviations from the mean in the same direction, this will result in a positive 'r' value.

FI>EE Inbound 2012	- 0.17	+ 0.95	FI Tourism demand	Table 10
RU>EE Inbound	- 0.86	+ 0.55	EE Border counts	Table 11
FI>EE Inbound 2012, overnight	- 0.60	+ 0.96	FI Tourism Demand	Table 13
FI>EE Inbound 2012, same-day	+ 0.76	+ 0.25	EE Tourism Demand	Table 14
FI>EE Inbound, overnight (raw data)	-	+ 0.98	EE Tourism Supply	Table 17
FI>EE Inbound, overnight (corrected)	+ 0.37	+ 0.98	EE Tourism Supply	Table 16
EU27>EE Inbound, overnight trips	+ 0.44	+ 0.98	EE Tourism Supply	Table 19
RU>EE Inbound, overnight trips	+ 0,.42	+ 0.92	EE Tourism Supply	Table 20
DE>EE Inbound, overnight trips	+ 0.62	+ 0.97	EE Tourism Supply	Table 20
LV>EE Inbound. overnight trips	+ 0.79	+ 0.74	EE Tourism Supply	Table 20
SE>EE Inbound, overnight trips	+ 0.53	+ 0.94	EE Tourism Supply	Table 20
UK > EE Inbound, overnight trips	+ 0.54	+ 0.93	EE Tourism Supply	Table 20
LT>EE Inbound overnight trips	+ 0.47	+ 0.92	EE Tourism Supply	Table 20
US>EE Inbound overnight trips	+ 0.31	+ 0.93	EE Tourism Supply	Table 20
FR>EE Inbound, overnight trips	+ 0.40	+ 0.97	EE Tourism Supply	Table 20
EU27>EE Inbound, nights spent	+ 0.66	+ 0.96	EE Tourism Supply	Table 21
RU>EE Inbound, nights spent	+ 0.70	+ 0.93	EE Tourism Supply	Table 22
DE>EE Inbound, nights spent	+ 1.05	+ 0.94	EE Tourism Supply	Table 22
LV>EE Inbound. nights spent	+ 1.37	+ 0.74	EE Tourism Supply	Table 22
SE>EE Inbound, nights spent	+ 0.85	+ 0.86	EE Tourism Supply	Table 22
UK > EE Inbound, nights spent	+ 0.91	+ 0.91	EE Tourism Supply	Table 22
LT>EE Inbound nights spent	+ 0.99	+ 0.91	EE Tourism Supply	Table 22
US>EE Inbound nights spent	+ 0.81	+ 0.91	EE Tourism Supply	Table 22
FR>EE Inbound, nights spent	+ 0.90	+ 0.90	EE Tourism Supply	Table 22
EU27>Harju County	+ 0.45	+ 0.98	EE Tourism Supply	Table 23
EU27>Pärnu County	+ 0.65	+ 0.99	EE Tourism Supply	Table 23
EU27>Tartu County	+ 0.86	+ 0.88	EE Tourism Supply	Table 23
EU27>Saare County	+ 0.66	+ 0.97	EE Tourism Supply	Table 23
EU27>Lääne County	+ 0.77	+ 0.94	EE Tourism Supply	Table 23
EU27>Lääne Viru County	+ 1.08	+ 0.94	EE Tourism Supply	Table 23

Table 10: Selected coherence indicators for outbound and domestic tourism activities (from Report 3b).

Segment covered	Coverage (Avg. CC2AB)	Consistency (r)	Reference statistics	Report 3b
EE>EU27, 2012	+ 0.57	+ 0.70	EE Tourism Demand	Table 24
EE>FI Outbound	+ 0.31	+ 0.75	FI Border Interview	Table 25
EE>FI Outbound	+ 0.20	+ 0.97	EE Border Interview	
EE>EU27	+1.00	+1.00	EU27 Tourism Supply	Table 26
EE>FI Outbound	+ 1.41	+ 0.89	FI Tourism Supply	Table 27
EE>LV Outbound	+ 0.84	+ 0.98	LV Tourism Supply	Table 27
EE>SE Outbound	+ 1.50	- 0.83	SE Tourism Supply	Table 27

EE>LT Outbound	+ 0.62	+ 0.88	LT Tourism Supply	Table 27
EE>DE Outbound	+ 0.80	+ 0.99	DE Tourism Supply	Table 27
EE>FR Outbound	+ 1.33	+ 1.00	FR Tourism Supply	Table 27
EE>PL Outbound	+ 0.44	+ 0.77	PL Tourism Supply	Table 27
EE>UK Outbound	+ 0.55	+ 0.86	UK Tourism Supply	Table 27
EE: Domestic, overnight trips	N/A	+ 0.98	EE Tourism Supply	Table 28
EE Domestic, overnight trips	N/A	+ 0.79	EE Tourism Demand	Table 29
EE Domestic, nights spent	N/A	+ 0.91	EE Tourism Supply	Table 28
EE Domestic, nights spent, Estonia	N/A	+ 0.82	EE Tourism Demand	Table 30
EE Domestic, nights spent, Harju county	N/A	+ 0.68	EE Tourism Demand	Table 30
EE Domestic, nights spent, Ida-Viru	N/A	- 0.23	EE Tourism Demand	Table 30
EE Domestic, nights spent, Lääne county	N/A	+ 0.88	EE Tourism Demand	Table 30
EE Domestic, nights spent, Lääne-Viru	N/A	+ 0.65	EE Tourism Demand	Table 30
EE Domestic, nights spent, Saare county	N/A	+ 0.79	EE Tourism Demand	Table 30
EE Domestic, nights spent, Tartu county	N/A	+ 0.50	EE Tourism Demand	Table 30
EE Domestic, nights spent, remaining EE	N/A	+ 0.64	EE Tourism Demand	Table 30
EE Domestic, same-day-trips	N/A	N/A	N/A	

Report 3b provides reasonable explanations for most of the deviations between statistics based upon mobile positioning data and the respective reference statistic. Evidently, the reference statistics have their own methodological boundaries, as has mobile positioning data (as has been outlined in much detail in Report 3a). Therefore, blaming deviations between the two methodologies only on a lack of validity and reliability in the mobile positioning data gathering and processing is not always justified. However, as error margins in traditional sources are relatively well known, they should be used as a valid reference point, as has been carried out here.

Furthermore, it has to be taken into account that some of the statistics shown above have been computed based upon corrected mobile positioning data. The correction has usually been based upon accommodation statistics for the relevant country (Estonia, in this case).

The high rate of consistency in variation over time, however, makes the use of mobile positioning data as an additional source for tourism statistics quiet promising.

3.1.6 Resolution

3.1.6.1 Spatial Resolution

In terms of spatial resolution, mobile positioning data can be delivered at the granulation of space being covered by one antenna (cell), as long as the *cell_id*/*CGI* (*Cell Global Identity*) and/or the geographical coordinates are available (see Chapter 4 of Report 2,

specifically Sections 4.1.7 and 2.1.2 of Report 3a). The area covered by a single antenna varies, depending on the population density (see Figure 3 for an example). Using estimates for probabilistic geographical distribution can even deliver finer granulations (Report 3a, p. 15).

However, such a fine granulation goes beyond the requirements of European tourism statistics. Report 3a states: 'This is not the requirement in Regulation (EU) No. 692/2011. According to the official regulation, estimates on NUTS level 2 and by type of locality are the most detailed level required for the supply-side statistics and the country level for the demand-side statistics'.

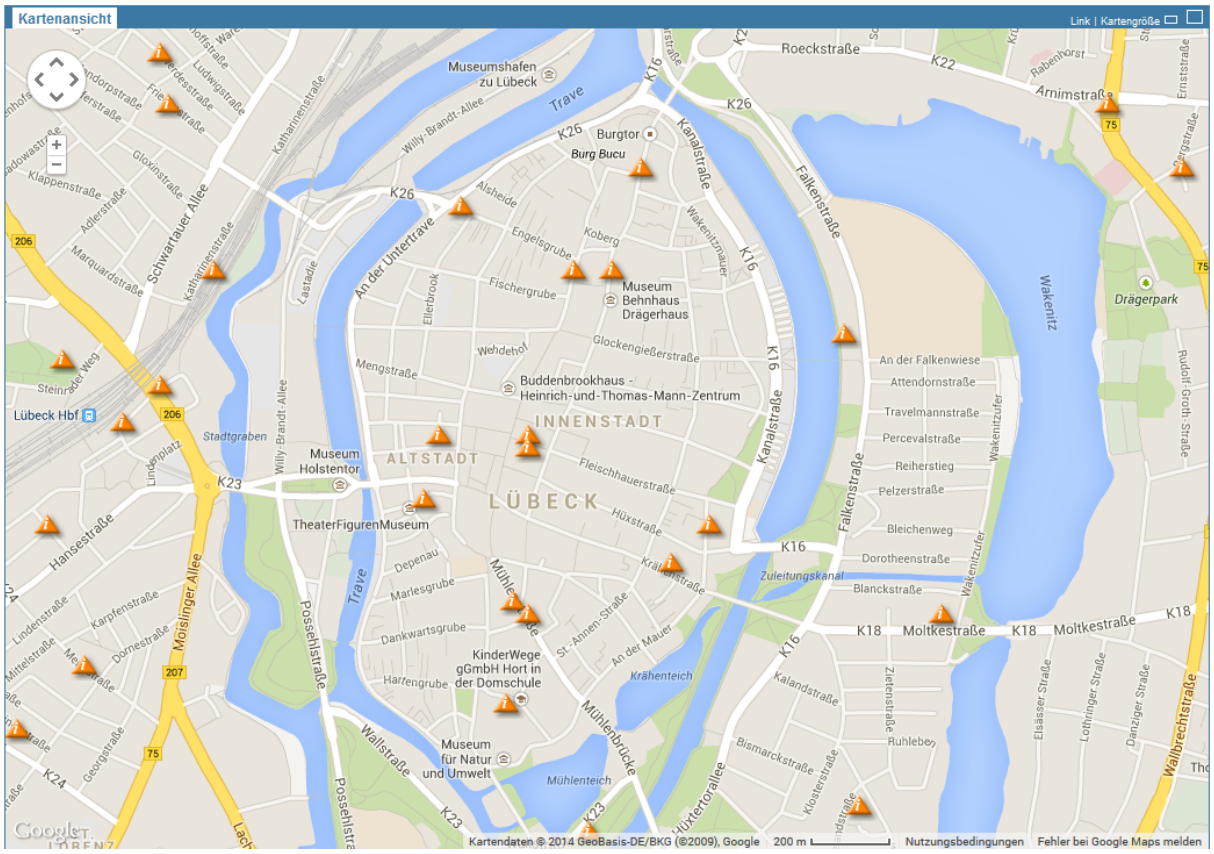

Figure 3: An example showing antennae distribution in a densely populated area (the old town area of Lübeck in Germany), Source: Bundesnetzagenutr/German Federal Network Agency; orange triangles indicate antennae.

3.1.6.2 Time Resolution

The mobile positioning data which serves as a potential basis for tourism statistics always contains a time stamp of the event recorded. As the time stamp has a granulation of one second or less, this is theoretically the minimum time-slice to be derived from the data.

However, some limitations have to be taken into account. Usually, the records are event-based, meaning that a certain event, such as sending or receiving a text message or voice call, triggers the record process (see Section 4.1.2 in Report 2 for a list of triggers).

Therefore, the registries in question do not constantly take a snapshot of the mobile devices that have linked into the network, but rather only do so when an event triggers the recording. It has been shown in Report 2 that data from the CDR (Call Data Record) registry is the most easily obtained. Such data, however, relies mostly upon active triggers such as calling or texting. Therefore, the time granulation when using CDR data depends on the frequency of a user actively using the network (i.e. making or receiving a call or sending or receiving a text message).

The impact on the recorder length of stay/trip and the actual length of stay/trip is hard to assess and depends heavily on the usage of the mobile phone (e.g. tourists who send 'Hi, I have safely arrived' or receive the welcoming message from the network and 'Hi, I am now getting on the airplane' versus passive tourists who do not use the phone with such continuance). From the comparison of same-day and overnight visitors, the difference can be up to 35 % (see Report 3a, Section 3.2.1 and Report 3b, Section 4.4.4) meaning that a third of the same-day trips that are compiled from mobile positioning data are actually overnight trips. There is also a problem with trips that only include one event (single call or SMS during the trip) that generate a number of trips that might be very short visits or are just under representation of longer same-day or even overnight trips. Such quality problems should be handled in the estimation process if there is a reference statistic concerning the distribution of the lengths of the trips (Figure 4).

There are methods for a more reliable detection of the presence of the mobile phone in the network for domestic and inbound roaming data, but not for outbound data (see Report 2, Section 4.1.2).

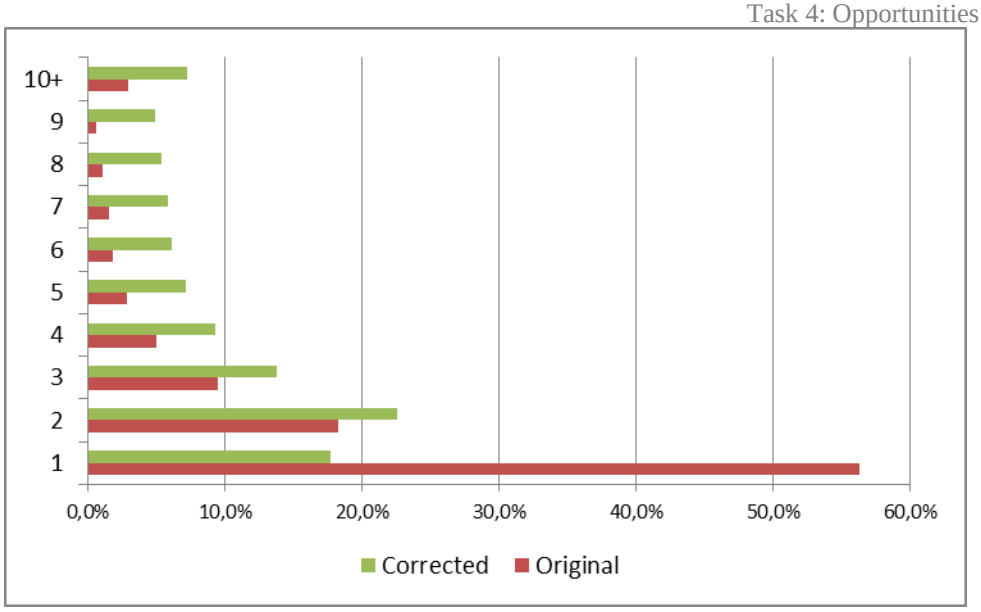

Figure 4: The difference between corrected and uncorrected proportions of trips that are based upon duration in days.

3.1.6.3 Conclusion

Within the limitations of the ability to determine the 'usual environment' (domestic tourism), the spatial and timely granulation of mobile positioning data is much finer than required by the regulation and opens up new possibilities when it comes to the use of tourism statistics.

Concerning the aspect of the length of trips, the comparison based upon reference data (the proportion between the duration of actual trips) has to be made, along with possible corrections (extending the duration of trips) to trips that contain only one event, as well as the distribution of trips which have different durations.

3.1.7 Conclusion (Quality Dimension)

The assessment dimension 'quality' shows some aspects where mobile positioning data can be superior to traditional sources of data for tourism statistics. In other aspects of quality, however, no advantages can be found in using mobile positioning data:

- Completeness: it has been shown that mobile positioning data cannot replace traditional sources because mobile positioning data does not deliver complete coverage of any sector that is relevant in tourism statistics as covered in Regulation (EU) 692/2011. However, mobile positioning data might very well deliver additional indicators, as discussed in Section 3.3 below;
- Timeliness: if work processes could be fully integrated and automated, it could be expected that mobile positioning data becomes available much sooner than data from

other sources. This perspective, however, is theoretical. A more realistic scenario would make use of mobile positioning data as quick indicators for specific sectors;

- Validity: in terms of validity, no specific advantages can be found in using mobile positioning data;
- Accuracy: when compared to surveys, a number of advantages can be shown when it comes to using mobile positioning data (smaller sampling error, no memory gaps). However, in terms of accuracy, the mobile positioning data-based definition of 'usual environment' in measuring domestic tourism flows is an issue. It is, however, also problematic in surveys where people themselves need to say whether they ventured outside of their usual environment or not. Mobile positioning data would allow the use of a very strict definition of the concept, such as a trip outside an individual's own county or LAU which allows a more harmonised use of the concept throughout the EU;
- Consistency: there is, on average, a high grade of consistency in variation over time between statistics produced on the basis of mobile positioning data and the respective reference statistics. Although both over-rating and under-rating can occur, consistency in variation over time is more important when it comes to the implementation of additional information or quick indicators;
- Resolution: apart from the problem of determining the 'usual environment', mobile positioning data can be more finely granulated than required, both in terms of time and space, which would open up new possibilities when it comes to the use of tourism statistics.

As a consequence from the assessment of the quality dimensions with regard to official tourism statistics, it seems impossible to replace traditional data sources by mobile positioning data. However, mobile positioning data can deliver additional information in several ways, being suitable for the following areas:

- Delivering key indicators of tourism statistics in a more timely manner than today (quick indicators);
- Delivering data in a finer spatial resolution than is possible (or required) today;
- Calibrating deficiencies in other sources simply by having another data source available. This, however, is dependent upon whether the bias can be evaluated or not. If it can, the bias-corrected estimates can in theory be used for calibration purposes.

3.2 Cost Dimension

This chapter assesses the cost/burden dimension in terms of using mobile positioning data for tourism statistics. The cost/burden dimension will be assessed from the three perspectives outlined in Table 11.

While the tourist's and industry's perspectives are being addressed briefly, the focus of this chapter is on the offices who produce official statistics, specifically Eurostat and the National Statistical Institutes (NSI).

Table 11. Cost/burden aspects to be assessed.

Cost/burden aspect	Description
Tourists	What costs/burdens will tourists have to bear when provide mobile positioning data and how can these costs/burdens be compared to today's situation? Can the burden in terms of interview time be reduced through the introduction of mobile positioning data usage (less interviews, less questions)?
Industries (reporting units)	What costs/burdens can be expected for relevant industries (specifically MNOs) when providing mobile positioning data and how can these costs/burdens be compared to the costs/burdens of industries delivering data today (specifically accommodation establishments)?
Statistical offices	What costs are being incurred for the statistical offices at subnational, national and European levels in order to collect, process and deploy mobile positioning data and how can these costs be compared to today's situation?

For the cost dimension in Statistical Offices, the Generic Statistical Business Process Model (GSBPM) can be used as a framework for assessment.

3.2.1 Tourist's Burden

Using mobile positioning data does not incur any direct cost or burden to the users of mobile devices because the original data is being produced independently from their future use in statistics and without users even noticing.

This latter fact, however, might be an indirect burden for users: giving away data on their whereabouts, possibly without them even noticing. This, on the other hand, is a question of privacy regulation, as extensively explained in the report on Report 2.

Furthermore, it needs to be taken into account that the use of mobile positioning data as a complimentary data source does not increase and in some cases (such as in the replacement of some indicators) can reduce the burden on tourists (see also the Methodological Manual on tourism statistics, Version 2.1, 2013, Chapter 3.2.3.1 for a discussion on the methodological impact of respondent burden):

- Each year, thousands of interviews are being conducted within the EU for the purposes of gathering data for the demand-side of tourism statistics;
- In many countries, accommodation establishments are required to have their guests fill in some paperwork which can be used as a basis for statistical reporting.

It is not effective to make an extensive analysis of these factors here. It can be concluded that using mobile positioning data will certainly not increase the burden on tourists, but does have the potential to reduce it. In the complete picture of any assessment, this will, however, have only a minor impact.

3.2.2 Industry Burden

Mobile network operators (MNOs) are the main reporting units. Therefore, it is the cost that is incurred by the MNO which is the main point of relevance here.

This cost has been analysed extensively in Report 2 (Section 5.1). The main results for a sample MNO are shown in Table 12.

Table 12. Cost for a sample MNO with 10 million subscribers, including inbound, domestic and outbound (from Report 2, Section 5.1).

Task/phase	max cost (EUR)	min cost (EUR)
Implementation: pilot phase	135,000	63,000
Implementation: data extraction	551,000	106,000
System Maintenance, annual cost	158,000	51,000

For a typical configuration with a country that has a population of around sixteen million people, three relevant MNOs (with ten, five and one million subscribers respectively) and average cost structures, implementation costs of around EUR 1.1 million, and annual maintenance costs of EUR 320,000 are being assumed for the maximum extent of reporting work (processing that is carried out by MNOs); and implementation costs of around EUR 210,000 and annual maintenance costs of EUR 100,000 are being assumed for the minimum extent of reporting work (processing that is carried out by the NSI).

Obviously, a possible reduction of costs when replacing traditional sources with mobile positioning data also has to be taken into account. Assuming that today's accommodation statistics are being produced by collecting monthly reports from accommodation establishments (above the threshold), it can be concluded that a considerable burden is currently being imposed on the accommodation industry (even when taking into account the fact that in some countries the reporting of arrivals needs to be carried out for the police or immigration authorities). For example, the average time used for reporting on

accommodation statistics is estimated to be thirty minutes (Estonia, STATEST), and in Germany this is forty minutes per month, or 485 minutes per year (DIW 2006).

Eurostat's statistics on accommodation establishments shows an estimate of more than 500,000 establishments in the EU28, thereby providing an average of almost 20,000 establishments per country that are reporting monthly on various aspects, as requested in Regulation (EU) 692/2011. This, of course, is only true if all establishments need to report and not only a sample is being interviewed.

3.2.3 Statistical Offices Burden

The *Generic Statistical Business Process Model* (UNECE 2013) is a reference model, against which actual business processes can be mapped (Figure 5). The GSBPM should be sufficiently flexible to apply in all of the statistical processing scenarios, including (with the latest Version 5.0) against the growing use of non-survey sources like mobile positioning data. As such, the model provides a good basis for cost estimations.

The model presented here includes only activities and expenses by National Statistical Institutes (NSI) and does not include any expenses made by mobile network operators (see previous section).

Because foreseeable activities depend on the setup of the system, two scenarios are presented here. The cost scenarios presented here cover the costs of

- a full processing chain by MNOs (*minimises* the workload and costs on the NSI side, and are therefore noted as 'min');
- only extraction and preparation of the data by MNOs (*maximises* the workload and costs on the NSI side, and are therefore noted as 'max').

The latter basically means that the core of the processes is shifted from the MNO to the receiving party (e.g. an NSI). Obviously, workload and costs are higher for the NSI in the 'maximum' scenario, but the total cost for the production is lower (on the account of not having separate systems for each MNO).

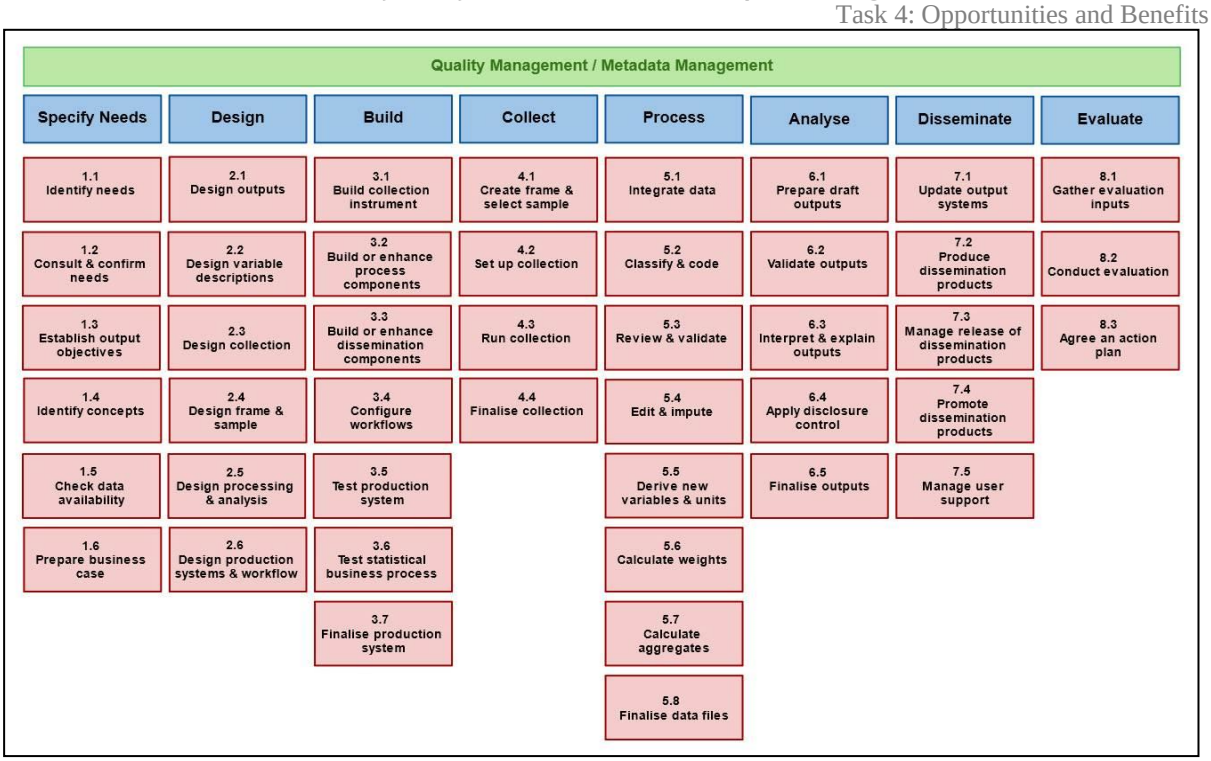

Figure 5: The 2008 *Generic Statistical Business Process Model* (Source: UNECE 2013).

It must be clearly stated that all figures produced in this chapter are estimates based upon averaged model assumptions. The concrete costs of implementation and maintenance will vary from country to country.

Table 13 summarises the workload deducted from the previous sections. The detailed GSBPM model is presented in Annex 2.

Table 13. An estimate of costs incurred in all phases (summary). Estimations are based upon a country with a population of around 16 million, with three MNOs (large, medium and small).

	Estimated time (days)	Estimated time (implementation, days)		Estimated time (data updates, days/year)	
Adopted sub-process	Pilot	Max	Min	Max	Min
Phase 1 - Specify Needs	26	15	15	0	0
Phase 2 - Design	40	67	47	0	0
Phase 3 - Build	85	1,385	260	0	0
Phase 4 - Collect	29	13	0	0	0
Phase 5 - Process	75	60	60	36	12
Phase 6 - Analyse	30	0	0	78	84
Phase 7 - Disseminate	0	0	0	24	24
Phase 8 - Evaluate	70	0	0	0	0
Quality Management	0	0	0	18	36
Metadata Management	0	0	0	12	12
TOTAL (days)	**355**	**1,540**	**382**	**168**	**168**
Hardware/IT maintenance (EUR)	20,000	250,000	25,000	50,000	5,000

TOTAL (EUR)	**126,500**	**712,000**	**139,600**	**100,400**	**55,400**

3.2.3.1 Cost Examples

Obviously, a cost assessment only makes sense when comparing data for a new solution against data for solutions that are already in place. Table 14 shows some key indicators being explained below.

Table 14: Annual man-days spent on processing, data validation, making corrections, and the imputation of the data in mobile positioning, demand and supply-side surveys (monthly updates). * Finnish demand survey data collection shared with the Consumer Survey.

	Mobile positioning data		Status Quo			
			EE: Demand-side	EE: Supply-side	FI: Demand-side	FI: Supply-side
	Max	Min				
Country size	16 million		1.3 million		5.4 million	
Data scope	Inbound, outbound, domestic		Outbound, domestic	Inbound, domestic	Outbound, domestic	Inbound, domestic
Update period	Monthly		Quarterly	Monthly	Annual	Monthly
Collection method	Automated data collection		Survey	Data request	Survey	Automated data collection
Annual cost (man days)	264	168	600		690*	

Burden for Statistics Finland (Demand Statistics) - status quo

Approximately 28,000 Finnish residents are included in the sample for the Finnish Travel survey. The survey is conducted together with the Consumer Survey in a joint telephone interview covering both surveys. There is also an additional survey for participation in tourism activities which is conducted in January and concerns the previous year as a whole.

The direct resources necessary for conducting the survey are roughly 5,000 man-hours on an annual basis. All steps in the data collection process are shared with the Consumer Survey, apart from the interview itself. In addition, up to 500 man-hours per year are required for processing the data, including data validation, correction and imputation.

Currently the full burden for data collection and processing is only three man-years due to joint data collection duties that are shared with the Consumer Survey. A standalone Finnish Travel Survey would nearly double these resources. This estimate doesn't include data analysis and dissemination which requires roughly half a man-year.

Potential for a mixed-mode data collection (a demand survey augmented with mobile positioning data)

The current Finnish sample size of 28,000 results in roughly 17,000 responses, unit non-response being 39% for 2012. While this sample is more than enough for high-level

classifications (e.g. purpose of trip, type of accommodation used, domestic or outbound) it quickly falls short when breaking down the trips by destination. In case of Finland, less than 30 destination countries have enough observations for dissemination. Increasing the sample size is not feasible and would not significantly improve the accuracy in case of rare destinations.

Mobile positioning data can potentially augment the current demand surveys by providing an administrative data source for trips by destination. In such a scenario, the number and duration of trips would be based upon mobile positioning data while tourism expenditure and ratios (purpose of trip, type of accommodation, means of transport, etc.) are based upon demand survey. The sample size of the demand survey could be decreased considerably since the survey doesn't need to support breakdown by destination, reducing the 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.

Estonian Demand-side survey

According to the Estonian tourism demand survey, the most time consuming phases of the survey are the data collection and processing phase, which consist of up to 64% of all hours spent on the survey. Efficiency in the most time-consuming phases would provide a significant decrease of burden for NSIs. However, it is difficult to see how the use of mobile positioning data could improve efficiency without lowering the quality of demand statistics. Although the mobile positioning data does not provide the same level of detail as the tourism demand survey, it enables the production of more timely estimates of the total numbers of trips and nights spent, but for estimates of the detailed tourism demand breakdowns, still a relatively large sample survey is needed for producing reliable estimates, although probably with a smaller sample size compared to toady’s situation.

Table 15: An estimate of costs incurred in all phases in the Estonian Demand Survey, 2013 (summary); Source: Statistics Estonia.

Adopted sub-process	Current demand survey, Estonia (days)
Phase 1 - Specify Needs	5
Phase 2 - Design	40
Phase 3 - Build	40
Phase 4 - Collect	120
Phase 5 - Process	270
Phase 6 - Analyse	80
Phase 7 - Disseminate	10
Phase 8 - Evaluate	5

Quality Management	15
Metadata Management	15
TOTAL (days)	**600**

3.2.4 Conclusion (Cost Dimension)

The total cost of the use of mobile positioning data depends largely on the size of the country and the number and size of MNOs, and also the timeliness of the process, specifically in terms of the shorter retrieving cycles. As has been mentioned, there are two basic setup options available, depending on the allocation of the main processing practice – either for the MNOs or for the NSI. The total cost of the system is smaller in the latter case; however, the direct cost for the NSI is larger (see Figure 6). When MNOs handle processing duties, the total cost is higher, but the cost for the NSI is smaller. The main benefit for the allocation of processing to the NSI is the ability to analyse sensitive micro-data and the cost of changing the methodology, because it is much more difficult (read expensive) to change the processes in the systems of several MNOs as opposed to one system in the NSI. The following examples are based upon a country which has a population of around sixteen million inhabitants, with three MNOs that have ten million, five million and one million subscribers respectively. The implementation and production costs are different for each MNO.

Table 16. Level of burden for MNOs and the NSI (for a country with a population of about 16 million, covering three MNOs, with 10, 5 and 1 million subscribers) when MNOs process the data (EUR).

	Pilot	**Implementation**	**Production**
MNO #1	135,000	551,000	158,000
MNO #2	95,000	386,000	111,000
MNO #3	41,000	165,000	47,000
NSI	127,000	140,000	55,000
Burden for MNOs	**270,000**	**1,102,000**	**316,000**
Cost for the NSI	**127,000**	**140,000**	**55,000**

Table 17. Level of burden for MNOs and the NSI (for a country with a population of about 16 million, covering three MNOs with 10, 5 and 1 million subscribers) when the NSI process the data (EUR).

	Pilot	**Implementation**	**Production**
MNO #1	63,000	106,000	51,000
MNO #2	44,000	74,000	36,000
MNO #3	19,000	32,000	15,000
NSI	190,000	712,000	100,000
Burden for MNOs	126,000	211,000	102,000
Cost for the NSI	190,000	712,000	100,000

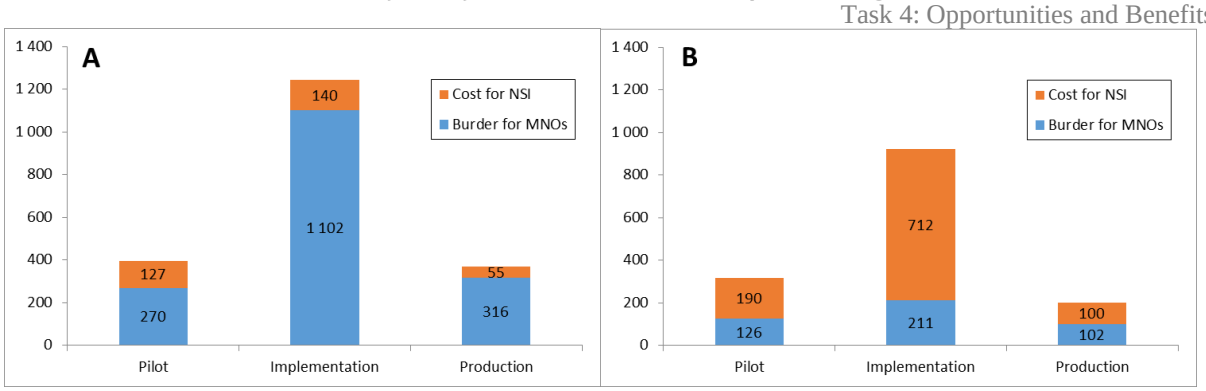

Figure 6. A graphical representation of the burden on MNOs and costs for NSIs according to the processing allocation for the data (A - processing carried out by MNOs; B - processing carried out by the NSI) (figures show thousands EUR).

The main cost of the system lies in the implementation part. The running costs are relatively low because of the automation potential of the system - the data is digitally collected and processed and the only part for which manual work has to be carried out is validation, quality assurance and the interpretation of the data. It is calculated that for the model country with 16 million inhabitants, annual workload amounts to 264 days per year, for the demand survey in Estonia the work amounts to 600 and in Finland 690 days.

Table 18. Benefits and constraints in cost dimension.

Minimum extent - processing carried out by MNOs, processed aggregated data provided to NSIs.	**Maximum extent - MNOs provide extracted raw data to NSIs, who handle full processing of the data.**
Benefits • The cost to the NSI is much lower when compared to the alternative (if the NSI does not have to reimburse the expenditure of MNOs); • Smaller processing complexity for the NSI. The NSI already receives aggregated results data;	Benefits • The cost of the overall system is lower (the burden is on MNOs + cost of the system for NSI); • Methodological changes incorporated more easily;
Constraints • The cost of the overall system is higher (the burden is on MNOs + cost of the system for the NSI); • More difficult (and expensive) to implement methodological changes;	Constraints • The cost to the NSI is much higher when compared to the alternative; • Higher processing complexity for the NSI;

There is potential for efficiency increase in tourism demand statistics by decreasing the survey samples of Demand Surveys. Currently, the highest contributor to cost in surveys is the collection and processing – 64% of total man-hours in the Estonian case. In a mixed-mode data collection, where overall number of trips and nights spent with breakdowns by country of origin are based on mobile positioning data, survey sample sizes could be

decreased. By calculation, a 30% decrease in the collection and processing efforts leads to a 17% decrease in the overall annual man-days needed for the demand survey. The running costs of mobile positioning data could be offset by the decrease in survey cost. However, the overall cost-benefit analysis is dependent on the specific country.

3.3 Access Dimension

In this chapter, only the access to new indicators through the usage of mobile positioning data will be discussed. From the list of indicators structured by type, this chapter focuses on new/additional indicators (Table 19), either in existing tourism statistics sectors or even in new ones.

Table 19. Types of indicators.

Types of indicators	Description
Complimentary indicators/alternative data source	Mobile positioning data delivering the same information for tourism statistics when compared to indicators that are available today
Replacement indicators for reasons of efficiency	Mobile positioning data possibly replacing indicators for tourism statistics that are available today (same output), but with less costs (less input)
Replacement indicators for reasons of effectiveness	Mobile positioning data possibly replacing indicators for tourism statistics that are available today with comparable costs (same input), but with better quality (more output)
New/additional indicators	**Indicators that are not available today for tourism statistics**

Within the chapter, two main fields are being addressed:

1. New indicators for official tourism statistics, both within the coverage of Regulation (EU) 692/2011 (i.e. on a European level) and outside the regulation (i.e. on a national level);
2. New indicators for tourism applications outside the field of official tourism statistics (at both the EU and national level).

3.3.1 New Indicators for Official Tourism Statistics

As outlined in Chapter 2.1 of this report (quality assessment), mobile positioning data can be particularly useful as a source for additional indicators within the official tourism statistics, an additional source for calibration for traditional sources, and as a quick indicator for tourism statistics in relevant fields. In certain fields, also a replacement of more traditional data sources might be taken into account. These aspects will be further detailed in this chapter.

Tourism statistics cover a relatively wide range of aspects. In addition to those addressed in Regulation (EU) 692/2011 (mainly accommodation statistics and statistics tourism demand), some countries produce additional statistics with relevance for tourism.

The following list shows the statistics currently being produced in the European Statistical System as a whole or by parts of it, although it makes no claim to be complete (see also the Methodological Manual for tourism statistics, Eurostat 2013b, p. 230 and the project Report 3b, Chapter 3.1).

Table 20. Tourism statistics, within and without Regulation (EU) 692/2011 (which makes no claim to be complete).

Statistics	**Scope**	**Data sources**
Accommodation statistics (Re.g. 692/2011, Annex 1, Section 1-2)	Internal tourism activities in accommodation establishments above the threshold	complete counts (census) or incomplete counts (sample) based upon declaration by accommodation establishments
Accommodation statistics (Re.g. 692/2011, Annex 1, Section 3)	Internal tourism activities in non-rented accommodation	usually sample-based household-surveys
Statistics on participation, tourism visitors and tourism trips, overnight and same-day (Re.g. 692/2011, Annex 2)	National tourism activities	usually sample-based household interviews
Statistics produced on the basis of border interviews	Inbound visitors to the reference countries	usually sample-based surveys
Statistics produced on the basis of border control counts	Inbound visitors to the reference countries	usually complete counts
Statistics produced on the basis of passenger counts	Passengers in the process of using the respective transport mode, such as air, sea or train	usually either complete counts based upon industry declarations and/or sample-based surveys;
Statistics produced on the basis of reports of travel agencies	Customers of travel agencies participating in national tourism activities	usually (sample-based) industry reporting
Payment card data	Transactions of payment cards within the reference country	Accounting settlement systems
Tourism Satellite Accounts	Contribution of tourism to national income	Usually several data sources[10] including those 'that are not usually part of current NSI statistical programmes'[11]

3.3.1.1 Mobile Positioning Data as an additional Reference

It is clear that the statistics indicated (Table 19) are subject to the specific methodological boundaries imposed by the respective data collection and processing mechanisms. Therefore, in certain scenarios it is possible to derive additional information from mobile positioning data which can be used as an additional source of reference.

[10] Cf.. UN/Eurostat/WTO/OECD (2010).
[11] Cf. Eurostat (2009/2010).

These include:

1. Reference information on cross-border mobility: mobile positioning data is particularly reliable when the 'country of origin' information contained in the dataset can be analysed (see the examples given in Report 3b);
2. The information on participation, tourism visitors and tourism trips, overnight and same-day (Annex 2 of the official regulation) is based upon survey data, with a sample size of several thousand interviews per year or wave, usually just big enough to allow a reporting on NUTS 2 level. As opposed to this, mobile positioning data produces much larger datasets which, as a general rule and provided that there is no systematic bias, provides more precise information on objects within the scope of the data. It is clear that at the same time the usage of larger datasets (provided they are unbiased) provides not only finer granulation but also smaller corridors of confidence to the data applied, thereby making it possible to use such data as an additional source of reference.

However, if the coverage bias is assumed to be stable over time, then mobile positioning data can still be used, if not for level estimates then at least for change or trend estimates.

3.3.1.2 Mobile Positioning Data as a Source for Replacement Indicators

The example of cross-border travel between Finland and Estonia (see Report 3b) shows that in a well-defined framework, mobile positioning data can quite reliably reproduce the results that are obtained from other sources (cf. Section 3.1.5 in this report).

Knowing that in the given example Statistics Finland has ceased to conduct border interviews for cost reasons, the potential of mobile positioning data to replace this sort of data sources in cross-border travel becomes evident (see Section 3.4.1 of this report for a more extensive description of this case).

3.3.1.3 Mobile Positioning Data as a Source for Additional Indicators

Additional indicators for official tourism statistics from mobile positioning data can be expected in the following fields:

1. As accommodation statistics are not covering establishments below the ten bed threshold, a more or less important part of tourism cannot be reflected. In rural areas, this under-coverage can reach 50% and more. If mobile positioning data could provide an estimate of the total number of days spent in an area and the number of nights spent

in accommodation establishments which have ten beds, where more is known from tourism statistics, the missing number can easily be computed and thereby adding an additional indicator to the tourism statistics system. Whether this indicator can be seen as additional or replacement can be disputed. Assumed that the requirements of the official regulation are the base reference at this point in time, any information on the complete tourism flows can be considered 'additional', because it is outside the scope of the official regulation as it is today.

2. This would also help with the calculation of total tourism supply volumes (including the volume of accommodation below the threshold and accommodation without charge).
3. The same argument can be drawn for the demand-side, because a wider scope becomes possible. At the moment, only members of the population who are aged fifteen and over are observed (because interviewing children is not allowed). Mobile positioning data covers also children with mobile phones, and a recent study in Germany has shown the considerable size of the segment of children who are travelling unaccompanied by their parents or grandparents (and who are therefore usually not being reported in surveys) (NIT/DWIF/iconkids (in publication)).
4. Longitudinal/tracking analysis: all of the traditional data sources serve as cross-section data collections: a specific subject is always being observed at a given point in time - be it as a user of an accommodation establishment, a traveller crossing a border, a customer in a travel agency, or as a visitor during one or many visits within their trip. Mobile positioning data, in theory, can allow for a tracking analysis, providing information on one subject over time, including the individual's origin, destination and all of the tourist touch points in between. This allows for a completely new form of tourism statistical analysis which is based upon origin-destination relationships and the recurrence of visits. All this, however, is only possible if a tracking using a single mobile device ID is possible for a certain time. Cut-off times in ID of several hours or even one day obviously do not allow this. Provided that a unique ID can be followed over the length of the tourism trip, the 'customer journey' in its real form could be remodelled.

3.3.1.4 Mobile Positioning Data as a Source for Quick Indicators

As opposed to other, more traditional data sources, mobile positioning data can be collected and processed rather speedy. The reasons for this are that only a very limited number of sources have to be contacted in order to compile the data, and that all of the

relevant data is available in an electronic format (as opposed to paper questionnaires that are possibly still in use, such as in terms of accommodation statistics, border interviews or household interviews) and, provided that the system runs for a certain length of time, a high level of automation can be achieved (cf. also Section 2.5 of this report).

Assessing the concept of 'quick indicators' means to distinguish between two different aspects of 'quick':

1. Firstly, it needs to be evaluated whether a better timeliness can be achieved. Timeliness is the time lag between the end of a reporting period and the time the statistical report about this period is being published (see Section 3.1.2 of this report for a more formal discussion of timeliness).
2. Secondly, if timeliness can be improved, it can be asked whether a higher frequency of reporting will be possible and sensible. The frequency of reporting means that data is not being published monthly (as is standard for tourism supply-side statistics) or annually (as is standard for tourism demand-side statistics on a European level, although some countries like Estonia publish such data quarterly), but in shorter time intervals, like biweekly, weekly, daily or even shorter.

As to timeliness, today's time lag between the end of the reporting period and the date upon which complete details about this period are being published can typically be assumed to be about four to eight weeks for supply-side statistics and four to six months for survey-based demand-side statistics even on the national level (and the time lag can be considerably bigger on a European level). Something that needs to be taken into account is the fact that the first published supply-side data is sometimes not final, but is instead a series of rather good estimates that is based upon imputations for those establishments that do not report on time.

These time lags can be considerably reduced by using mobile positioning data once the transmission and processing procedures are in place. We can safely assume that the data that is needed for quick indicator figures can be drawn from the MNO database within a few days of the end of the reporting period. Processing time can also be reduced to a few days as well, so that it is realistic to be able to expect some quick indicators within a week or, at a maximum, two weeks after the end of the reporting period. Obviously, these indicators are preliminary in terms of absolute level, seasonal corrections, or the identification of usual environment, etc. However, what these figures can do is to show development over time - and

they can do that quite reliably, as Report 3b has shown. Therefore, in a sense, for these figures it is less important how big the estimation error is, as long as it is constant over time.

For inbound travellers by country (today covered by supply-side statistics), the authors see the following indicators to be specifically suitable as quick indicators:

- Number of inbound tourists and trips;
- Their length of stay;
- Their destinations.

For internal tourism activities (today covered by supply-side statistics), monthly flash indicators could include (within the methodological boundaries given in Reports 2, 3a and 3b):

- Number of domestic tourists and trips;
- Number of visits to specific points of interest (domestic and inbound tourism activities).

For national tourism activities (today reflected in demand-side statistics), monthly flash indicators could include:

- Number of domestic and outbound tourists and trips;
- Length of stay of domestic and outbound trips.

The aforementioned topics mainly address timeliness. At a secondary level, we can ask if the reporting period can be shortened and therefore, whether mobile positioning data can help not only to make data availability faster, but also to deliver data more frequently.

Technically, almost near-real time automated data processing can be established (with a delay of up to one hour only). The aspects that affect the cost/usefulness of the speed of the delivery of the data are the following:

- The faster the delivery of the indicators, the costlier is the implementation and the maintenance cost of the system (majority SLA - Service Level Agreements to react very fast on any technical problems of the system);
- Not all data appears immediately in the systems of MNOs - some data might be stored only after several hours or, in some cases, days. Therefore the quick indicators are often not fully representing the actual situation or 100% of the data that would be finally collected for the specific time. There is an option (and this is most probable) that the indicators are updated regularly but not for

the most up-to-date data -the updates are provided every day for data that is up to two days old (as an example, the indicators for the period leading up to 13 May are presented on 15 May);

- Because the recalculation process (identifying the country of residence, usual environment, duration of the trips, etc.) requires a longer period of data for specific subscribers, quick indicators differ when compared to final reliable outcomes. The difference between initial results (quick indicators) and the final outcome is described in Report 3a Section 2.8.1;
- An important part of the data processing routine is the quality assurance and possible corrections in the estimates, which is something that is carried out on a partially manual basis. In the case of a near-real time data feed, a considerable level of work is involved if it is carried out every time the data is updated.

The cost of the implementation and annual maintenance of the quick data processing system when compared to the monthly updates can be up to 3 and 4 times higher for daily updates and 6 and 9 time higher for near-real time updates (see Report 3a, Section 2.8.1).

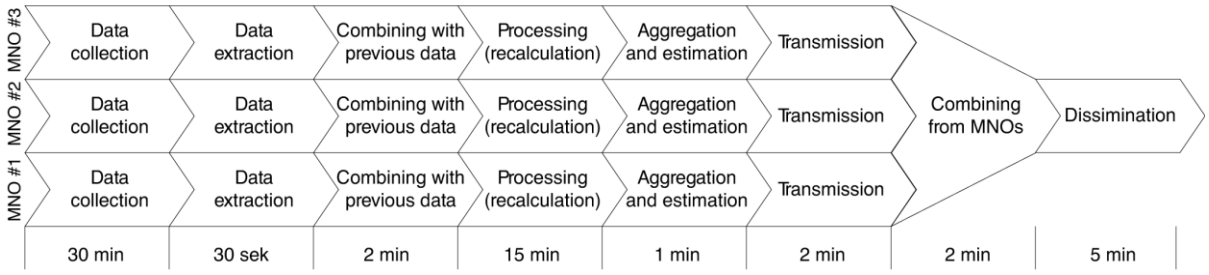

Figure 7: Data processing sections which are important from the point of view of timeliness. An example showing one hour quick indicators.

Because of the recalculation process, confirmed results and unchanged final indicators are provided for any data that is no longer recalculated, i.e. after six or twelve months (depending on the longest recalculation criteria - e.g. identifying the country of reference). This might depend on the recalculation specifics - e.g. the recalculation and correction of the results can be omitted if the identification of the specific identifiers is only retrospective and the calculation takes only for the new data but such an approach may result in temporarily shifted or even incorrect results.

Two illustrations are provided in Figure 8 and Figure 9. The student example shows a university town in which out-of-town students arrive at the university town in September and

leave in May-June. In Figure 8 (with no recalculation) one can observe the increase in the number of tourists in September when students are arriving at the university. Logically, those students in September can be considered as residents as they plan to stay in the town for the next eight or nine months. If no recalculation is carried out, then the number of residents only increases in February/March when, based upon the six month retrospective calculation, those students are identified as residents. If a recalculation is carried out, then although until the February/March the picture is similar to when the data is recalculated in February/March, the picture changes drastically and those students are identified as residents six months earlier.

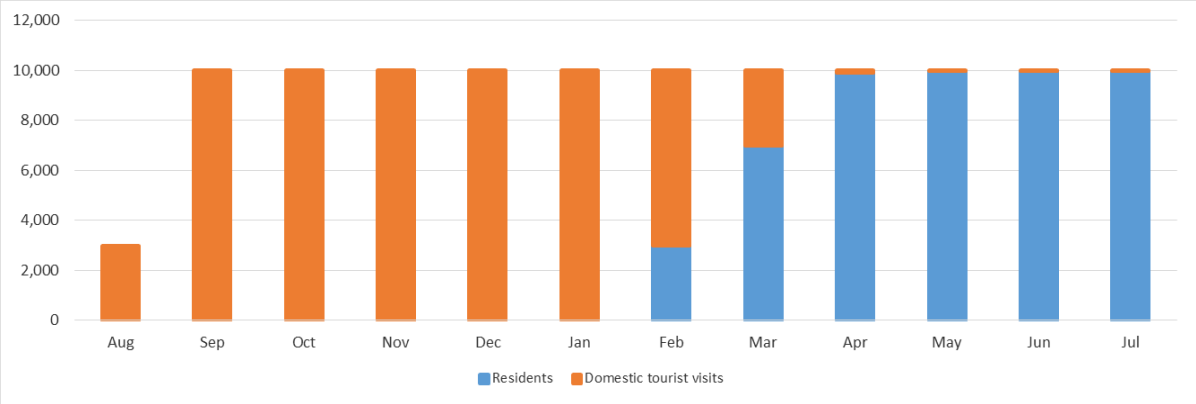

Figure 8: Residents and domestic tourists without recalculation. Identification of the place of residence (and usual environment) based upon a retrospective six month window.

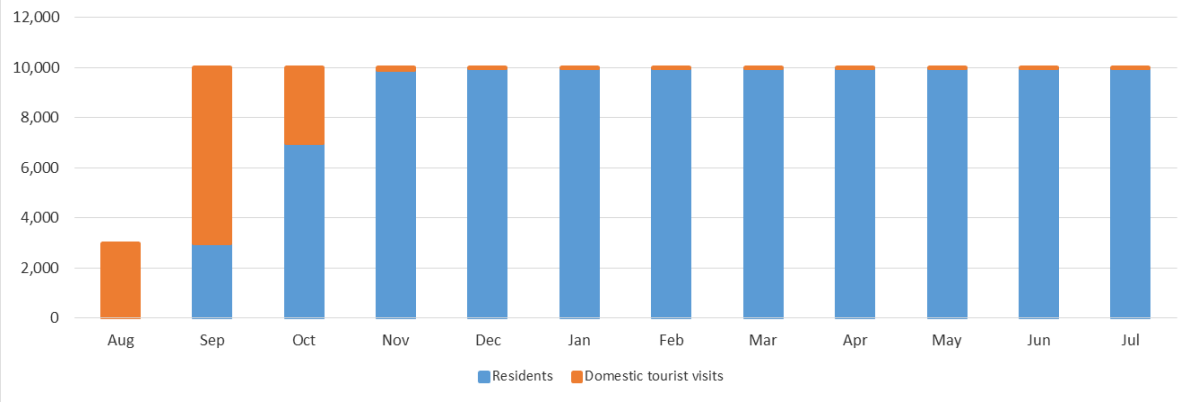

Figure 9: Residents and domestic tourists with recalculation. Identification of the place of residence (and usual environment) based upon a retrospective six month window.

When considering that indicators are incorrect or shifted as long as the final reach of the recalculation is achieved, all such indicators can be considered ‘quick’ as opposed to ‘final’. However, some measures can be taken to decrease such problems specifically with foreseeable phenomena (e.g., the increase in students as tourists can be foreseen to some extent).

Another approach to definition is the presence of manual quality assurance. It is neither reasonable nor possible to conduct manual quality control on the results in near-real time (1 hour to 24 hours). This manual control by itself takes much longer. So basically all indicators that are created within 24 hours, can be labelled 'uncontrolled' or 'quick' as opposed to 'confirmed'.

Independent of the use of mentioned measures, basically all indicators can be used as quick indicators with the notion that these might be corrected during the period of recalculation and/or during the manual quality assurance. But some indicators are more affected by the recalculation due to the nature of the tourism form they represent. Inbound and outbound tourism indicators are more 'resistant' to future recalculation. Domestic indicators are more inclined to change after recalculation (e.g. because of the criteria for usual environment). There is no inclination for the quality assurance - all data can be erroneous.

3.3.2 New Indicators for Tourism Outside Official Tourism Statistics

Apart from contributing insights to official tourism statistics, mobile positioning data has the potential to provide a wide choice of indicators for whole new fields of application within tourism activities:

1. They can be used to analyse tourism in a much more *detailed spatial and temporal statistics* when compared to other methodologies used in today's tourism statistics;
2. As there are *none of the thresholds* that are available for accommodation statistics, the mobile positioning data can potentially be used to analyse tourism in non-paid accommodations and in establishments that fall below the threshold;
3. Longitudinal mobile positioning data can distinguish between *repeat and first-time visits*;
4. Combined with aggregated personal data for the subscribers, it would be possible to use the mobile positioning data to analyse the travel behaviour of different *segments*.

Table 21 is an overview of the potentially useful different kinds of indicators that are based upon mobile positioning data. The following sections provide some examples of how indicators that are based upon mobile positioning data are already being used or are planned to be used in these new fields of application.

Table 21. Possible fields of application for mobile positioning data in tourism-related fields.

Field of application	Possible indicators	Remarks, further explanations
Detailed *spatial* statistics	Number of visitors at a customised location	

	Share of different nationalities	
	Duration of stay [12]	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data
Detailed *temporal* statistics	Number of visitors during a customised period	Such as defining day visits/multi-day visits
	Share of different nationalities	
	Duration of stay	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data
Detailed *spatiotemporal* statistics (e.g. event monitoring)	Number of visitors at a customised location during a customised period	Such as event monitoring
	Share of different nationalities	
	Duration of stay	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data
No thresholds	Accommodation: defining and measuring different segments (summer house tourism, non-paid accommodation)	Algorithms needed to define different segments
	Demand-side: at the moment the population observed in demand-side is citizens aged 15 and over.	Mobile positioning data does not have age limits.
Longitudinal data	Repeat/first-time visits	
Segment data	Residence, sex, age, family status, etc.	Requires subscribers data

3.3.2.1 More Detailed Spatial Statistics

Czech example

Due to the insufficiency of existing data when it comes to tourism flows in the Czech Republic at a geographic detail, the Czech Tourist Authority - Czech Tourism - has decided to look for new ways of monitoring tourism activities, ways which are innovative, cost-efficient and effective.

Czech Tourism started to use the mobile positioning data for monitoring tourism flows in April 2012. The data is collected for a total of 45 locations which have high tourism potential (UNESCO sites, mountain regions, spa towns). The combination with the existing data provides a comprehensive view of the number of visitors to the monitored locations, the

[12] Same-day and overnight trips. The regulation states that domestic tourism same-day visits need to be collected for the first time in 2016 to cover the reference year 2015. For outbound tourism, information on same-day trips is to be collected for the first time in 2015 (to cover the reference year 2014). The domestic same-day estimates which are based upon mobile positioning data are given in Task 3b, Paragraph 4.7.

country of origin (in the case of domestic visitors the region of residence), and the seasonal nature of tourism flows during the year. This is the relevant tourism data that is essential for developing tourism strategies and marketing planning. The data was delivered to Czech Tourism as daily aggregated data covering the number of domestic and international visitors. Figure 10 shows an example of the Excel file for domestic visitors at one UNESCO site.

Datum	Návštěvníci	Návštěva	Lokalita	Měsíc	Kategorie	Den	Set2	Set3	Týden	Rok-Měsíc	Kvartál
01.06.2012	Domácí	354	Telč	červen	UNESCO	pátek	pracovní den	pracovní den	23	2012-06	Q2
02.06.2012	Domácí	662	Telč	červen	UNESCO	sobota	víkendový den	sobota	23	2012-06	Q2
03.06.2012	Domácí	629	Telč	červen	UNESCO	neděle	víkendový den	neděle	23	2012-06	Q2
04.06.2012	Domácí	384	Telč	červen	UNESCO	pondělí	pracovní den	pracovní den	24	2012-06	Q2
05.06.2012	Domácí	318	Telč	červen	UNESCO	úterý	pracovní den	pracovní den	24	2012-06	Q2
06.06.2012	Domácí	322	Telč	červen	UNESCO	středa	pracovní den	pracovní den	24	2012-06	Q2
07.06.2012	Domácí	576	Telč	červen	UNESCO	čtvrtek	pracovní den	pracovní den	24	2012-06	Q2

Figure 10. Aggregated daily volume data (Source: Czech Tourism).

The project was terminated in May 2013. The data was provided to the destination managers who were able to use it for their strategic planning. From the very beginning, the project was focused on testing the new method and on being able to alert tourism professionals to the new possibilities that were being made available in tourism monitoring (see also Report 1, Usage Case 3)

French example

Spatiotemporal analysis of foreign tourists in Paris based upon anonymous signal data from a mobile network (Orange). GSM technology can capture large amounts of spatiotemporal data. The trajectories that are inferred from such data provides additional information for analysing human mobility. A conceptual data model allowing the spatiotemporal dimensions of GSM data to be modelled is created, and a first instance of this model is presented: the spatial behaviour of foreign tourists and the estimation of visits for the most important Points of Interest in the Paris region (see also Report 1, Usage Case 5).

Belgian example

As a small and mostly rural region, the tourism ministry of the German-speaking part of Belgium cannot work with the actual official statistics on the spatial level that they need to work with. They also feel that accommodation establishments tend not to report their guests/nights spent accurately. They would need a good and at the same time affordable source of data in order to be able to measure and understand tourism where it happens (e.g. hiking in the forests) in order to plan and evaluate infrastructure investments. Further challenges that cannot be addressed with today's official tourism statistics are day tourism, being a border region (BE/DE/NL) with lots of cross border travel. For all of their challenges they see the future potentials of indicators based upon mobile positioning data.

3.3.2.2 More Detailed Temporal Statistics

Tallinn-Helsinki (twin cities) example

Human mobility study of movements between Tallinn and Helsinki (twin cities) was conducted in 2012 based upon mobile network data (Silm, Ahas & Tiru 2012). Due to the fact that the state borders in the European Union have been opened, there is a lack of border statistics. Most of the traffic between the two countries takes place through a few harbours and airports located in the capitals under study. Although data from shipping companies allows for the assessment of traveller numbers and economic statistics in order to evaluate the number of tourists staying overnight, there are many transit and one-day visitors between the two cities, which is why mobile network data was employed. The problems with the data arise from having to rely on one operator to estimate the entire traffic flow, and also from people not once using their phones abroad during the day.

The data showed that the majority of visits Estonians make from Estonia to Finland last for one day. This data could be analysed temporally, to arrive at day visitors visiting Finland any month, season, year. In addition to day visitors, other visit lengths were given (from one day up to 366 days). The same data was given for Finns visiting Estonia. The behaviour was observably different. For example, a distinct seasonality was detected among Finnish one-day visitors to Estonia, whereas no such seasonality was detected for Estonians visiting Finland. (See also Report 1, Usage Case 8).

3.3.2.3 More Detailed Spatio-temporal Statistics

Event monitoring (Flanders example)

VisitFlanders would like to monitor the big events in Flanders that are coming up in 2014 and 2015 in Flanders to monitor the WWI remembrance. All of these events can be defined by a distinct spatial area and time. Established methodologies are not feasible or too expensive (e.g. how to measure the volume and structure of a 90km people-torch-line).

In order to find out if it is possible to work with mobile phone positioning data VisitFlanders talking to the relevant MNOs. Apparently they can offer ready-made aggregated data that is based upon positioning data and subscribers' information, e.g. volume in a defined area at a given time, origin of the visitor (nationality and Belgian province), route to event, duration of stay, single/partner/family, repeat/first time visitors. VisitFlanders is planning to put the work out to tender for the relevant MNOs and to start-up the project in spring 2014.

Event Monitoring (Estonian example)

The research on visitors of large events like Madonna, Metallica and others presents the possibility to investigate the visitors of the events on very small spatial and temporal form. Analysis of the consistence of the visitors provides possibility to study where the visitors come from (foreign or domestic tourist), when did they arrive to the event (and the city where the event takes place), when do they leave, what are other places that are visited, how many of them are first-time visitors and how many are repeat visits, have foreigners visited the country and the location of the event before the event and/or have they visited later after the event (repeating, loyal tourists). Also a cluster analysis on the similar events can be made to find the similar events (visited by same people). The effect on local and national economy can be estimated based upon the characteristics of visitors of each event (e.g. the number of spent days/nights by visitors of the event) (see also Report 1, Usage Case 7).

3.3.2.4 No Thresholds

Estonian example

The seasonal variability of population was studied in Estonia, and a methodology was developed for the monitoring of the short-term mobility of population with mobile positioning data (Silm & Ahas 2010). The locations of calculated home anchor points of telephone users were analysed by use of the dataset from EMT, Estonia's biggest mobile operator, over a period of two years. The results showed that approximately 5% of the population of Estonia change their place of residence seasonally. The number of residents rises during the summer months in coastal areas, the surroundings of cities, and in specific `dacha' areas. Most of these seasonal migrants originated from cities and towns. The timing and geography of the seasonal migration patterns studied showed the different directions and causes of seasonal moves. The methodology developed for the monitoring of short-term migration is suitable for the monitoring of movements over more extensive territory. In order to determine the causes and composition of these migrations, however, one must use additional survey studies or observations.

3.3.2.5 Repeat Visits

Estonian example

Loyalty-based segments and profile of concert-tourists in Estonia are investigated in University of Tartu. Importance of events in tourism has increased in recent decades. Events are important because they are diverse in their function: help to reduce seasonality, bring different types of tourists to destinations, add new opportunities to discover the destination.

Specific event's catchment area and indicator 'repeat visit' is new value added by passive mobile positioning method. Traditional methods (accommodation statistics, border statistics) don't include information about repeat visits and attractions visited, events attended in destination.

Aim of this study is to figure out different loyalty-based segments in events' visitors and evaluate importance of events in generating tourism in Estonia. For that we use data from six concerts in Estonia. The results show that concert visitors spend more time in Estonia than regular visitors and they are more likely repeat visitors than regular visitors. Organising events in Tallinn helps to market Tallinn as a destination, brings here new segments of tourists and stimulates repeat visits of tourist from nearby countries (see also Report 1, Usage Case 7).

Frequent visitors between Estonia and Finland.

The cross-border commuting study between Helsinki and Tallinn mapped well the transnational lifestyle, even when such studies are a great challenge in the open-border EU. Because the ferry is a very common transport mode between the two Baltic Sea capitals, shipping data provides the number of passengers crossing the sea. Mobile positioning data was used to analyse the behaviour of the visitors. As a result, it could be determined that the relative importance of Estonians who visit Finland 5 or more times a year is 19% of all Estonians visiting Finland and 22% for travellers who live in Tallinn. At the same time, the relative importance of Finns who visit Estonia frequently (5 or more times per year) is 4%. (see also Report 1, Usage Case 8).

3.3.2.6 Segment Analysis

'Smart Step' example

Smart Steps should help retailers, local councils, sporting arenas, major event organisers, public safety bodies, and other organisations to measure and understand the number of people visiting an area. Smart Steps is a joint initiative with GfK, Telefónica Dynamic Insights' market research partner, and is distributed by both companies.

Smart Steps uses anonymised and aggregated mobile network data to provide extrapolated trends about footfall by time, gender, and age.

Smart Steps shows how many people visit an area by time, gender and age. Organisations will be able to analyse the movements of crowds at any given place by hour,

day, week or month, and make like-for-like comparisons by area, as well as understand catchment patterns.

Retailers can use Smart Steps to understand footfall across their estate, tailor product promotions in existing stores, determine the best locations and formats for new stores, and measure the impact of marketing activity on footfall (see also Report 1, Usage Case 24).

3.3.3 Conclusion (Access Dimension)

It has been shown that mobile positioning data can contribute to the establishment of new indicators in tourism statistics, as well as in tourism applications that are outside official statistics.

Within the statistical system, the following benefits and opportunities could be found:

- Mobile positioning data is specifically interesting when it serves as quick indicators because of its superiority in terms of timeliness and its high coherence with reference statistics over time. Mobile positioning data would allow key indicators to be delivered at a rate that is much faster than today (specifically, but not exclusively, for inbound tourism), having the potential of reducing time lags from several weeks or months down to several days. In addition, also reporting frequency can be raised, so that reporting could no longer take place on a monthly or annual basis, but rather weekly or even daily. This last procedure, however, needs to be carefully evaluated in every specific case because cost and workload will rise considerably while precision will decrease.
- Mobile positioning data could be used as additional data source at the estimation stage for other surveys. For example, in surveys of foreign visitors where quota samples or other nonprobability samples are used;
- Additional indicators can be derived from the fact that mobile positioning data tends to cover the whole field of tourism and therefore also tourism in accommodation establishments which fall below the threshold and in private, non-paid accommodation (supply-side), as well as for tourists below the age of fifteen (demand-side). Although mobile positioning data does not make a difference between paid and non-paid accommodation, in combination with accommodation statistics non-paid accommodation statistics can be estimated without any burden to businesses or individuals;
- Furthermore, longitudinal analysis will become possible at least in theory and would therefore add to the set of statistical information.

The main value of mobile positioning data, however, lays in those tourism-related fields that lie *outside* official tourism statistics:

- Statistics more detailed in terms of space;
- Statistics more detailed in terms of time;
- Volume and structural breakdowns (nationality) of big events and other related applications;
- Statistics describing accommodation not being covered at all in tourism statistics;
- Segmentation data that relies on subscriber master data.

For each of these fields, existing usage cases were found.

3.4 Synergies Dimension

In the synergies dimension we look more closely into the opportunities and chances that can be expected to make themselves available in terms of mobile positioning data being used for tourism statistics and at the same time for other domains of official statistics.

The main domains to be looked at are:

- ECONOMY AND FINANCE: Balance of Payments (travel item);
- TRANSPORT: Passenger transport statistics (road, rail, air, sea);
- POPULATION AND SOCIAL CONDITIONS: Labour Market (commuting), Population (migration);
- Other domains.

Large portions of the analytical aspects have already been collected in Report 3a (Section 4).

3.4.1 Economy and Finance

In Economy and Finance statistics, mostly the Balance of Payments is relevant for assessing synergies with tourism statistics.

As opposed to tourism statistics, the travel item of the Balance of Payments looks at the import and export of travel services and is being expressed in currency units (and not head counts). The travel item states the total expenditures of travellers residing in the country of

reference in foreign countries (import of services) and the total expenditure of foreign travellers in the country of reference (export of services).[13]

Statistically, 'tourism' is a subset of 'travel' and consequently, (tourism) 'visitors' is a subset of 'travellers' (see Report 2, Section 1.3 and Section 3.1 for more details on definitions in tourism statistics). Groups being travellers but not tourists (= tourism visitors) comprise:

- Travellers arriving for employments, including seasonal workers;
- Travellers not leaving their usual environment (e.g. border workers);
- Travellers staying for more than one year.

For identifying these categories of travellers, the follow-up of the same mobile phone during a long period (up to one year) is necessary.

However, these 'obstacles' for tourism, namely the problem to accurately delineate usual environment, does not apply to the travel item in the Balance of Payments. Methodologically, during the processing of the data several objectives can be achieved depending on the setup of the system. For example, although domestic tourism concentrates on travel outside the usual environment, the same process can be extended to identify the trips within the usual environment.

An example for Finland might reveal the possibilities that are available in this field: mobile positioning data is currently used as one data source for calculating the expenditure of foreign travellers in Finland (export of services). Up to 2012, the main data source used for calculating exports was border interviews that were conducted with foreign visitors when departing from Finland. Since the border interview survey was terminated in 2012, a temporary method was developed for calculating export expenditure for the Balance of Payments.

The underlying calculation for the exports is carried out by country, and the country specific calculations are summed up for total export per quarter. The formula used in the calculation is:

$$\sum export_{tq}^{i} = \sum \left[X_q^i * \alpha_{tq} * \beta_{tq}^i \right]$$

[13] 'Travel credits cover goods and services for one's own use or which can be given away when they have been acquired from an economy by non-residents during visits to that economy. Travel debits cover goods and services for one's own use or which can be given away when they have been acquired from other economies by residents during visits to those other economies.' Int. Monetary Fund: *Balance of Payments and International Investment Position Manual* (BPM6), 2009, Num. 10.86 ff., p. 166.

where

$export^{i}_{tq}$= export expenditure of country i in year t, quarter q,

X^{i}_{q} = export expenditure for country i in basic year (2012), quarter q,

α_{tq} = an index describing change of tourist spending and prices of commodities typically consumed by tourists in year t, quarter q

β^{i}_{tq} = an index describing change of tourism flows for country i year t, quarter q.

The basic year for expenditure is 2012, for which the latest border interview survey data is available. The main source for price changes is the so-called tourism price index, which is a subset of the consumer price index that itself consists of commodities that are typically consumed by tourists.

The primary data source used for tourism flows depends on country. Typical data sources include: Accommodation Statistics, Border Control Statistics and statistics that are based upon mobile positioning data.

Mobile positioning data is used as primary data source for Estonian inbound tourism flows to Finland for two main reasons. Firstly, the mobile positioning data is compiled and published quarterly by the Bank of Estonia. Secondly, Estonians mostly use accommodation which is not registered in the Finnish accommodation statistics.

The method described above is temporary and will be used for a few years only. The permanent method will include new data sources for tourism expenditure based upon statistics of payment cards. As new mobile positioning data sources (inbound or outbound) become available, these sources are planned to be integrated into the permanent method.

3.4.2 Transport and Mobility

The usability of mobile positioning data for transport statistics has been addressed in Report 3a.

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:

- A huge sample size is needed from household travel surveys, e.g. the description of 40,000 trips has provided estimate of O-D flows between NUTS2 zones with an accuracy of 20% only for the largest traffic flows (Armoogum & Madre 1998);

- Road-side surveys are conducted on too few points on the network (moreover, since the late 1990's, for national and international trips, the police in France often refuses to stop vehicles, considering that surveying the traffic is not their main duty).

Therefore, MNOs can provide less costly and much more accurate matrices from mobile positioning data.

However, there are two drawbacks:

- The mode of transport is unknown, except for very slow (walking or cycling) or very fast modes (e.g. high speed train or air transportation (when the phone is OFF between one airport location and another));
- The purpose of a trip can be derived only from the location (e.g. for a special event) and from the duration of stay at this location (e.g. all day long at a workplace for full-time workers).

3.4.3 Population and Social Conditions

In population statistics, the spatial distribution of population (living, working) and mobility aspects like commuting could be relevant here.

As opposed to census-based statistics, mobile positioning data will always lack accuracy and will not offer the required level of detail or the opportunity for breaking it down. Therefore, a replacement of traditional data sources in population statistics by mobile positioning data is hardly imaginable.

However, in terms of timeliness, quick and short-term indicators could possibly be derived from using mobile positioning data as an additional source of population information. Also although the accuracy cannot compete with accuracy of census and registries, the data is more ubiquitous in time and can provide overall indications concerning the commuting, migration and internal migration information. Information on commuting, migration and internal migration is a valuable source for regional planning including infrastructure and public transport development. Combining mobile positioning data and data from administrative registers allows an estimate to be made on commuting that is due to working and commuting that is due to studying.

3.4.4 International Data Harmonisation

Besides the official statistics dimension, the implementation of mobile data in several countries can provide benefits when it comes to combining international data. From the international tourism point of view, mobile data provides information on inbound flows of foreign tourists who are visiting the country of reference (only the country of residence of the guest), while outbound data shows the destinations as well as transit countries to the destination countries.

One country's outbound flow to a specific foreign country is that foreign country's inbound flow. Combining these datasets can provide value in the following areas:

1. The harmonisation of datasets between all countries where mobile data source are used (ideally these should match if the same methodology is used);
2. Getting a better picture of international trips by being able to see destination and transit countries.

An example of usage in the latter case would benefit monitoring of the travel flows between countries that present international origin-destination matrices.

3.4.5 Conclusion (Synergies Dimension)

Analysis has shown that specific opportunities can be found with regard to

- The Balance of Payments Statistics, specifically the travel item;
- Transport Statistics for additional calibration of origin-destination matrices;
- Migration and commuting statistics.

For all three fields of application, existing usage cases could be evaluated.

The implementation of the system is rather expensive (see Section 3.2); however if the system is implemented for several domains (tourism including BoP, transportation and population), the additional costs for adding processing components is relatively lower.

3.5 Transmission Dimension

Transmission of indicators can mean different aspects of data procurement along the processing chain. There are three different steps in transmissions (Figure 11):

- The (automated) transmission of data from MNO's registers to indicator calculation:
- The (automated) transmission of indicators to the statistical database of the NSI;

- The (automated) transmission of indicators from the NSIs to the Commission and to other dissemination centres, e.g. 'European Virtual Tourism Observatory' (VTO).

The indicator calculation is described in more detail in Report 2, Chapter 4. It should be noted that the NSI can either do the steps described as indicator calculation themselves or outsource it completely or in parts.

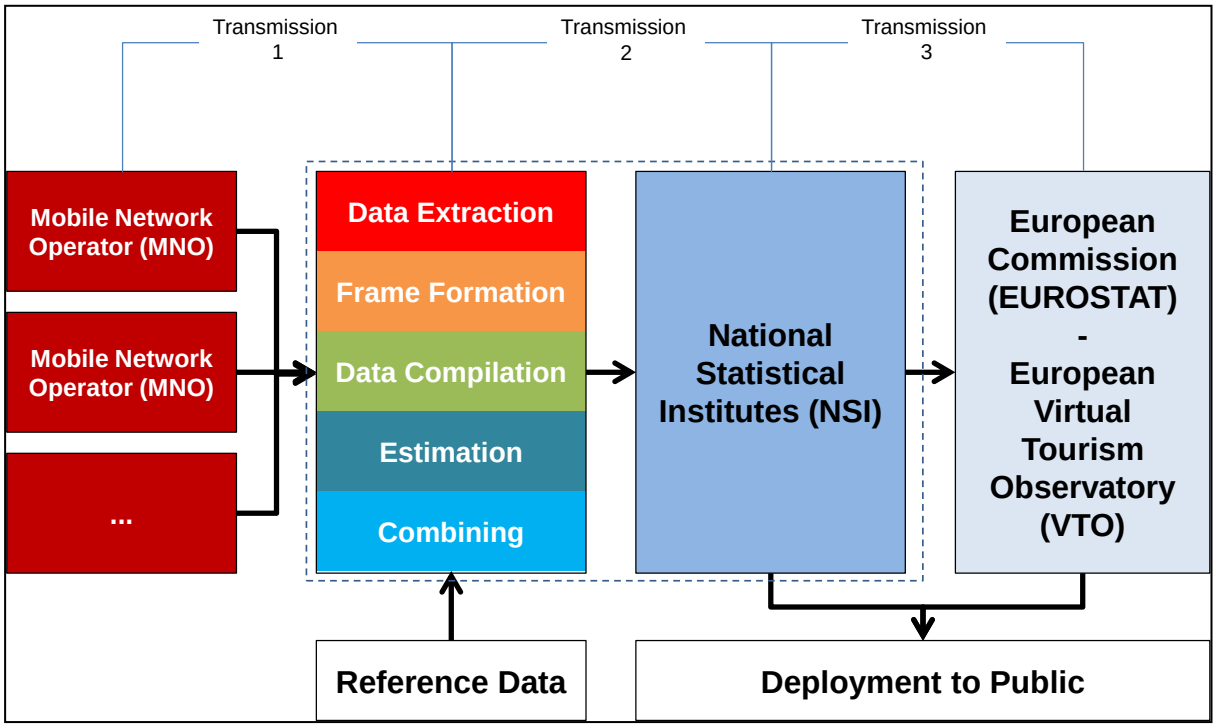

Figure 11: Three transmission steps.

3.5.1 Transmission from MNO's Registers

Depending on the setup for where the processing of the raw data to tourism indicators is conducted, there are two options for this stage of transmission:

- Internal extraction of the data from MNO's registers and transmission to the processing environment within MNOs (minimum extent);
- Internal extraction of the data from MNO's registers and transmission to the NSI outside for further processing environment within MNOs (maximum extent).

With both options, the initial extraction of the data should result in specifically-formatted and regulated prepared raw data which is ready for the further processing. This process is described in Reports 2 and 3a.

```
subscriber_id;country_id;event_time;cell_id
27436527823;FI;28.10.2012 12:53:38;24376243
27436527823;FI;28.10.2012 16:52:38;46243423
27436527823;FI;15.02.2013 22:03:08;24376243
27436527823;FI;17.02.2013 08:33:08;43523342
27436527823;FI;17.02.2013 19:45:18;93772342
```

Figure 12. An illustrative example of extracted ASCII raw data showing inbound roaming data from Report 3a.

```
cell_id;coordinates
24376243;55.92112 18.09451
46243423;56.11548 18.58742
43523342;56.12313 18.49385
93772342;56.33231 18.25124
```

Figure 13. An example showing cell reference data that is associated with the extracted data provided above.

This standard formatted data is either transmitted onwards to the next processing environment within the MNO or to the NSI. The specific data should be formatted according to agreed specifications (e.g. ASCII text file, CSV files, CDR ASCII format or binary format). The frequency of the extraction of the data (in other words, how regularly it is accessed) and the period that is covered by the extracted data is an important point to agree. For example, monthly data extraction by an MNO on the tenth day of every month with three separate files (including geographical references for antennae) for a period of one month between 1 January 2013 and 31 January 2013:

- inbound_20130101__20130131.txt;
- outbound_20130101__20130131.txt;
- domestic_20130101__20130131.txt;
- geographic_reference.txt.

3.5.2 Transmission to the NSI's Statistical Database

After the data is processed and the resulting data in aggregated form is ready, the process of transmitting the data to the NSI's statistical database is required. If MNOs are processing the data and transmit the result of the data to the NSI (minimum extent), the latter have to combine the data of all MNOs before storing it to the central statistical database (including any necessary quality assurance, validation procedures). In cases in which NSIs process the data, its storage in the statistical database is a matter of transmitting it from one system to another. The data model for the central statistical database is usually country-specific, but might be identical or similar to the standard proposed for transmission to Eurostat. The concrete execution of this transmission is largely dependent on the workflow being established between the NSI and possible service partners.

3.5.3 Transmission to the Commission and a VTO

Once the indicators are available for public deployment at the NSI, the general way of transmission of indicators to the Commission (Eurostat) will not differ a lot, if at all, from what is being done today (cf. Commission Implementing Regulation (EU) 1051/2011 and the Methodological Manual for Tourism Statistics, Eurostat 2013b, Section 2.9.1, 2.9.2, 3.7.1 and 3.7.2.). Because those indicators that are based upon mobile positioning differ from the official indicators required by Eurostat (e.g. 'Arrivals of residents at tourist accommodation establishments' - *occ_arr_res*), an updated transmission formats for specific indicators are required.

As to the transmission to a possible European Virtual Tourism Observatory (VTO), no transmission procedures have been agreed upon. The VTO is an initiative of the European Commission's Directorate-General 'Enterprise and Industry' (DG ENTR). The feasibility study has been awarded to the Tourism Intelligence Unit at the UK Office for National Statistics (ONS). Final results of the VTO project have not been published so far.

A discussion with DG ENTR revealed their specific interest in the following aspects of this feasibility study:

- Good practice with automated transmission of statistical data (case studies);
- General processes, technology and timing with automated transmission of statistical data;
- Quick indicators, near real-time and the involved cost.

The discussion also shows that at present both the implementations of a Virtual Tourism Observatory and of producing statistics that are based upon mobile positioning data are ongoing processes with some uncertainties in the details. Therefore, it is far too early to look for concrete connections between these two projects or even talk about technical details. Nevertheless, in the long run an inclusion of indicators in the VTO that are based upon mobile phone positioning data is seen as a potential benefit and should be kept in mind.

3.5.4 Conclusion (Transmission Dimension)

Fully automated transmission of the data plays a vital role in reducing manual work with the data and as a consequence getting gains in timeliness. In the case of mobile data, the structure of the data is very specific and a large number of errors can be identified automatically (e.g. missing values). The manual work has to be carried out in the validation, quality assessment and interpretation of the results, but this is fairly typical for any kind of

data source. It is suggested that all processes that can be automated, including transmission, should be automated, because when dealing with the large amounts of data, any manual processing is very time-consuming. In addition, manipulating data in a non-automated way is both costly and a potential source for processing errors.

Concerning the transmission standards, it is obvious that those should be set on the country-level, and follow as much as possible any kind of international standards (e.g. storing the data in the central statistical database according to Eurostat specifications). The regularity of the transmissions are a matter of technical setup, but it must be kept in mind that the timeliness of the data is heavily dependent on the cost of the system. Near-real time data transmissions and processing can provide great value to the users of the data, but are costly from the perspective of system implementation and maintenance cost.

4 Conclusions and Implications

This chapter brings together the main findings on the assessment dimensions and highlights the possible implications on the tourism statistics, both on European and on national level.

4.1 Conclusions

The analysis of has shown the following opportunities and benefits from using mobile positioning data in tourism statistics and in related fields:

- Because of the reduced number of data sources, mobile positioning data can be made available much quicker than data from traditional sources, once the necessary processes are in place. Therefore, an advantage in timeliness for selected **quick indicators** can be assumed. The authors think that inbound and outbound travel flows are particularly well suited for being reported based upon mobile positioning data.
- Another major advantage presented by mobile positioning data is that it allows users to gain **access to statistical information that was not previously available**. This is true for indicators within official tourism statistics (see above), but even more so for applications that are outside the current statistical framework. It has been shown that mobile positioning data can be used to estimate market sizes, provide fine resolution of data with regard to time and space, provide longitudinal

data/mobility characteristics and, in combination with subscriber data, a data enrichment source for segmentation.

- Altogether, at the moment we do not see a general scheme to replace existing indicators by those that are based upon mobile positioning data. This is true for supply-side statistics (accommodation statistics) and also for demand-side statistics. However, mobile positioning data can provide **additional information for both sources of tourism statistics**. One could even argue that if mobile positioning data could perfectly reflect domestic and inbound flows, the relevance of accommodation statistics would decrease because today it is the only source for data on inbound tourism.
- With regard to the statistical sectors that are covered by Regulation (EU) 692/2011, mobile positioning data can be useful at the estimation stage (e.g. in calibration). This is particularly true for inbound tourism, but also for domestic and outbound tourism, if suitable estimation procedures are in place. Therefore, mobile positioning data is a valuable additional source for **estimating market size** and development over time, for reasons that include its superior coverage for overnight trips when compared to supply statistics today.
- In addition to the requirements of the regulation, but still within the framework of official tourism statistics, mobile positioning data can deliver new information on those aspects, which is not required today mainly for methodological reasons. The two main fields of application are on the supply-side overnight tourism in accommodation establishments below the threshold and 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 finer than required by the regulation when using mobile positioning data and therefore open up new possibilities when it comes to tourism statistics.
- Accuracy: when compared to surveys, a number of advantages in using mobile positioning data can be shown (smaller sampling error, no memory gaps).
- In terms of coherence, mobile positioning data has shown that it can be consistent over time when compared to references statistics. This consistency in variation over time, however, is a most important prerequisite for quick indicators.
- Using mobile positioning data has the potential to reduce costs and burden, as well on the side of tourists (respondent burden in surveys) as on the side of

accommodation industry (monthly reporting). This potential, however, is limited to those cases in which mobile positioning data can replace other data sources. This today, as far as we can see, is not possible, and therefore additional costs seem to be the more probable development in the near future if and when mobile positioning data is used.

- Synergies can be found in the fields of transport statistics, migration statistics (commuting) and in the travel item for the Balance of Payments. It has been shown that mobile phone data can replace traditional sources (like border surveys).

4.2 Implications

At the moment, the implications for tourism statistics on European level, as governed by Eurostat and covered by Regulation (EU) 692/2011, are rather limited. The main reason is that the heterogeneity of rules and regulations concerning access to the data does not allow for a useful application in all Member Countries. Within the consortium, 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 that are being offered in the use of mobile positioning data for 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 that are tied to these implications even on a European level:

- The members of the European Statistical System (ESS), particularly the National Statistical Institutes (NSI), can profit from the large number of experiences and usage cases contained and analysed in this study;
- Eurostat can follow the development which will certainly take place in the coming years in selected Member Countries and therefore be prepared to implement the system when time has come;
- Eurostat, therefore, can function as a scout for application on national and regional level, as has been carried out on previous occasions (e.g. TSA, sustainability indicators).

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

1. Accessibility of data: as shown above, at the moment there is a broad range of regulations where privacy and access to personal data is concerned. It is up to the NSI to verify their country's position on the continuum from 'very liberal' to 'very restricted';
2. The specific situation of the sector 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 Finland and Estonia over the Gulf of Finland) can facilitate the assessment and implementation of mobile positioning data-based statistics. Also, geographical situations can exist which make it particularly hard to use such data (e.g. a high share of migrants producing roaming data without being part of tourism, as might be the case in Luxembourg).

4.3 Outlook: Internet Access and Satellite Positioning

All of the assessments in this Report (as in the previous reports) are mainly based upon the assumption that only a specific form of 'event data' from the network is being used for statistical purposes (see Report 2, Section 4.1 for a more detailed discussion). Basically, event data will at least be invoked when a speech or text call is made (producing a Call Detail Record, or CDR). If only CDRs are used, some limitations become inherent in the data. For example, in cases in which only one CDR event per subscriber is recorded, it will become impossible to estimate the length of stay - is it an overnight stay or a day-trip? Therefore, data quality is directly dependent upon the number of events (i.e. calls or text messages) that are produced by a subscriber.

It is obvious that the share of mobile internet users increased considerably during the last few years and will continue to do in the near future. Therefore, it can be expected that the usage rate of mobile internet access when travelling (or better: when being a tourist) will rise in the future as well and might even replace calls and texting (SMS). 'Always on' then means that a lot more 'event data' will be produced apart from the CDR and that a lot more event data can be processed in order to make granulation even finer.

A second aspect is that most mobile devices being sold today have the technical capability to identify its position using satellite positioning data (e.g. GPS). It is clear that only a fraction of today's subscribers receive GPS data constantly and allow applications on the mobile device to relay such data to third parties. However, there are some trend indicators that lead us to expect that satellite-based positioning will become more important in the next

few years. One is that even today major market research companies work with panellists who are giving away their mobility data in return for an incentive - not always, not everywhere, but often enough so that mobility studies can be derived from such information. Secondly, we can expect that location-based online services will become more important: telling you the way to the next pizza restaurant, revealing the rating of hotels within a two kilometre range of where you are now, or presenting additional Wikipedia information of the monument you are currently looking at through some sort of data glasses - all this requires the user to reveal his geographic position. Again, this technology is available today but its use is not accepted to a degree that would allow reliable statistics to be drawn from it. However, we would find it useful to closely monitor these aspects for the future.

References

Legal documents

Commission Implementing Regulation (EU) 1051/2011 of 20 October 2011 implementing Regulation (EU) No 692/2011 of the European Parliament and of the Council concerning European statistics on tourism, as regards the structure of the quality reports and the transmission of the data. Link: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:276:0013:0021:EN:PDF

Regulation (EU) 692/2011 of the European Parliament and of the Council of 6 July 2011 concerning European statistics on tourism and repealing Council Directive 95/57/EC. Link: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF

Regulation (EU) 99/2013 of the European Parliament and of the Council of 15 January 2013 on the European statistical programme 2013-17. Link: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:039:0012:0029:EN:PDF

Publications

Armoogum, J. & J.L. Madre (1998). *Redressement de l'enquête transports pour l'estimation de matrices Origine-Destination*. Rapport INRETS N° 223.

DIW (2006). Statistical Response Burden of Companies. *Weekly Report*, electronic edition 4. Berlin. Link: http://www.diw.de/documents/publikationen/73/diw_01.c.55648.de/diw_wr_2006-4.pdf

E-Plus Mobilfunk (2013). Mobilfunkmonitor 2013. Dusseldorf. Link: http://eplus-gruppe.de/wp-content/uploads/2013/06/e-plus_mobilfunkstudie_2013_final.pdf

European Statistical System (2009). ESS Standard for Quality Reports. Luxembourg. Link: http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf

European Statistical System (2012). Quality Assurance Framework of the European Statistical System. Version 1.1. Link: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/QAF_2012/EN/QAF_2012-EN.PDF

Eurostat (2009/2010). *Tourism Satellite Accounts in the European Union*. 4 Volumes. Luxembourg

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

Eurostat (2013b). Rolling Review Tourism Statistics: Executive Summary. March 2013. Link: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RRTourism%20Statistics%20Executive%20Summary.pdf

NIT/DWIF/iconkids (2014). Children and Youth Tourism in Germany, in publication, Kiel, Munich

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

Silm, S., R. Ahas, R. & M. Tiru (2012). Spatial Mobility between Tallinn and Helsinki in Mobile Positioning Datasets. Statistical overview. *Helsinki-Tallinn Transport and Planning Scenarios*. (Central Baltic INTERREG IV A Cross-Border Co-operation Programme). Tartu.

UN/Eurostat/WTO/OECD (2010). 2008 Tourism Satellite Account: Recommended Methodological Framework. (*Studies in Methods*, Series F, No. 80, Rev. 1). Luxembourg, Madrid, New York, Paris. Link: http://unstats.un.org/unsd/statcom/doc08/BG-TSA.pdf

UNECE (United Nations Economic Commission for Europe) (2013). 2008 Generic Statistical Business Process Model. Version 5.0. Geneva. Link: http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0

Abbreviations

CDR: Call Data Record

GSBPM: Generic Statistical Business Process Model

LAU: Local administrative unit

MNO: Mobile Network Operator

NSI: National Statistical Institute

NUTS: Nomenclature des unités territoriales statistiques

TSA: Tourism Satellite Account

VTO: Virtual Tourism Observatory

Annex 1: Use Cases by Scope

The following list sorts the usage cases as described in Report 1 into the field for the scoping matrix (see Figure 1 on page 9 of this report).

Tourism - official statistics

- Usage Case 2: Central Bank of Estonia - State Level Inbound and Outbound Statistics CHR

Tourism - further applications

- Usage Case 3: CzechTourism CHR
- Usage Case 4: O2 jetsetme.com
- Usage Case 5: Paris Tourism
- Usage Case 6: Tourist attraction centres
- Usage Case 7: Mass Events Monitoring
- Usage Case 29: TOURIAS Travel Guide

Outside tourism - official statistics

- Usage Case 1: Netherlands Mobility Statistics CHR

Outside tourism - further applications

- Usage Case 8: Helsinki-Tallinn International Commuting Study
- Usage Case 9: Israel travel study
- Usage Case 10: A1 Traffic Data Stream
- Usage Case 11: TomTom Traffic
- Usage Case 12: Estonian OD-Matrices
- Usage Case 13: Mobile Phone Positioning for Traffic Data Collection
- Usage Case 14: Deriving Origin-Destination Matrices
- Usage Case 15: TrendIT People Analytics
- Usage Case 16: Mobility Behaviour
- Usage Case 17: Longitudinal Mobility Study
- Usage Case 18: Mobility Behaviour (active, passive)
- Usage Case 19: Regional Commuting
- Usage Case 20: Fraunhofer GSM-GPS
- Usage Case 21: Graz in real time

- Usage Case 22: Big Data for Mobile Computing Research
- Usage Case 23: Mobility Patterns in Urban Sensing Data
- Usage Case 24: Telefonica Dynamic Insights
- Usage Case 25: Estonian Geomarketing Application
- Usage Case 26: O2, Telefonica Global Advertising Solution
- Usage Case 27: Sense Networks
- Usage Case 28: Mobility Behaviour and Social Networks
- Usage Case 30: LifeService112
- Usage Case 31: GIS-112

Annex 2. Generic Statistical Business Process Model Estimation for Mobile Positioning Data

		Estimated time (implementation, days)		Estimated time (data updates, days/year)	
Adopted sub-process	Pilot	Max	Min	Max	Min
Phase 1 - Specify Needs	26	15	15	0	0
1.1. Determine needs for information	1	0	0	0	0
1.2. Consult and confirm needs	1	0	0	0	0
1.3. Establish output objectives	1	0	0	0	0
1.4. Identify concepts	1	0	0	0	0
1.5. Check data availability	15	10	10	0	0
1.6. Prepare business case	7	5	5	0	0
Phase 2 - Design	40	67	47	0	0
2.1. Design outputs	3	0	0	0	0
2.2. Design variable descriptions	5	0	0	0	0
2.3. Design data collection	10	10	10	0	0
2.4. Design frame and sample	2	2	2	0	0
2.5. Design processing and analysis	10	15	15	0	0
2.6. Design production systems and workflow	10	40	20	0	0
Phase 3 - Build	85	1,385	260	0	0
3.1. Build data collection instrument	10	60	80	0	0
3.2. Build or enhance process components	40	1,200	100	0	0
3.3. Build or enhance dissemination components	10	40	20	0	0
3.4. Configure workflows	5	10	10	0	0
3.5. Test production system	20	30	20	0	0
3.6. Test statistical business process	0	15	10	0	0
3.7. Finalise production systems	0	30	20	0	0
Phase 4 - Collect	29	13	0	0	0
4.1. Create frame & select sample	20	10	0	0	0
4.2. Set up collection	3	1	0	0	0
4.3. Run collection	3	1	0	0	0
4.4. Finalise collection	3	1	0	0	0
Phase 5 - Process	75	60	60	96	12
5.1. Integrate data	0	0	0	0	0
5.2. Classify & code	0	0	0	0	0
5.3. Review & validate	10	0	0	12	0
5.4. Edit & impute	20	0	0	36	0
5.5. Derive new variables and units	15	0	0	0	0
5.6. Calculate weights	0	0	0	0	0
5.7. Calculate aggregates	30	60	60	48	12
5.8. Finalise data files	0	0	0	0	0
Phase 6 - Analyse	30	0	0	114	84

6.1. Prepare draft outputs	5	0	0	0	6
6.2. Validate outputs	10	0	0	72	36
6.3. Interpret & explain outputs	7	0	0	36	36
6.4. Apply disclosure control	5	0	0	0	0
6.5. Finalise outputs	3	0	0	6	6
Phase 7 - Disseminate	0	0	0	24	24
7.1. Update output systems	0	0	0	0	0
7.2. Produce dissemination products	0	0	0	12	12
7.3. Manage release of dissemination products	0	0	0	0	0
7.4. Promote dissemination products	0	0	0	0	0
7.5. Manage user support	0	0	0	12	12
Phase 8 - Evaluate	70	0	0	0	0
8.1. Gather evaluation inputs	10	0	0	0	0
8.2. Conduct evaluation	30	0	0	0	0
8.3. Agree an action plan	30	0	0	0	0
Quality Management	0	0	0	18	36
Metadata Management	0	0	0	12	12
TOTAL (days)	355	1,540	382	264	168
Hardware/IT maintenance	20,000	250,000	25,000	50,000	5,000
TOTAL Euro	126,500	712,000	139,600	129,200	55,400