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

Eurostat contract no. 30501.2012.001-2012.452

Report 2. Feasibility of Access

10 April 2014

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

Eurostat contract 30501.2012.001-2012.452

Report 2. Feasibility of Access

Final report

10 April 2014

Authors of this report (alphabetically):

Rein Ahas (University of Tartu), **Jimmy Armoogum** (IFSTTAR), **Siim Esko** (University of Tartu), **Maiki Ilves** (Statistics Estonia), **Epp Karus** (Statistics Estonia), **Jean-Loup Madre** (IFSTTAR), **Ossi Nurmi** (Statistics Finland), **Françoise Potier** (IFSTTAR), **Dirk Schmücker** (NIT), **Ulf Sonntag** (NIT), **Margus Tiru** (Positium, project coordinator)

Table of Contents

1. Introduction

Access to mobile positioning data from mobile network operators (MNOs) is the first challenge in the way of implementing such data in any domain. MNOs sell communications solutions (calls, information exchange) as their main business and they are not very forthcoming to provide their subscribers' sensitive data to third parties. The current study concentrates on the data from passive mobile positioning and no other methods (i.e. active mobile positioning) as explained in Section 4.1. The current report concentrates on questions of accessibility to such data from legal, business and technological points of view.

1.1. Aims, Content and Structure of Report 2

Report 2 assesses the potential opportunities and obstacles to gaining access to the passive mobile positioning data from mobile network operators. The main focus is the access to the data in order to produce official tourism statistics for national statistical institutes (NSI), but other usages are considered as well. The report concentrates on regulatory, economic and technological barriers along with practical access to the data. The knowledge for the current report is obtained from the experience of the consortium from the past projects, legal analysis of the subject, the practical process of accessing the pilot data from a number of operators or other organisations that have acquired the data and the experience of others gathered from the survey and interviews. The report consists of four main sections each concentrating on specific aspect of accessibility.

The section on the regulation and privacy protection aspects consists of the analysis of the legislation on EU and national levels (Estonia, Finland, France and Germany) that are relevant to the subject. The aim of the analysis is to assess the current situation and provide insight in current and future possibilities for accessing the data.

Technological aspects of the accessibility to the data are thoroughly discussed in the current report with the aim to help readers to understand the nature of the initial data that is in possession of MNOs and how this data should be extracted. This section is a description of the initial processes of the sequence of processes that continue in the Report 3a of the current study.

Financial and business-related aspects like the implementation and maintenance cost, business secrets of MNOs, effects of public opinion and benefits for MNOs are discussed in a separate section. This section provides insight mostly from the point of view of MNOs.

MNOs from consortium and other EU countries were contacted and introduced to the current project. In addition to MNOs, some organisations with access to the data have been contacted in order to clarify as many aspects in the data acquisition process as possible and asked if the data they possess can be used as a pilot data. Contacted MNOs have expressed interest in the project and see the potential in the usability of the data though many bring out concerns about how this can be done practically. The consortium was able to assess the characteristics and access conditions of the pilot data from Czech Republic, Austria, Netherlands, France, Germany, Belgium and Estonia. Because of the impossibility to get access to the data within the duration of the project, the cost thereof and/or the insufficient quality of the data, only the Estonian data was used in empirical tests in this study. The low quality of the data expressed in a variation of simple aggregation of subscribers per time unit without option for longitudinal calculation. The data from France and Germany required reimbursement and the cost of the data was too high.

During the compilation of the report, an online survey and interviews were carried out with experts and stakeholders whose input to the project presents a practical experience in the access and usability of the data as well as reflects the overall attitude from the stakeholders towards the use of mobile data in tourism and other domains.

The structure of this report consists of 7 sections and 8 annexes:

1) Introduction
2) Methodology of the Report
3) Privacy and Regulation-related Opportunities and Barriers
4) Technological Opportunities and Barriers
5) Financial and Business-related Opportunities and Barriers
6) Practical Experience on Accessing the Pilot Data
7) Conclusion

Annexes

1) Annex 1. List of Technical Abbreviations
2) Annex 2. EC Regulatory Documents
3) Annex 3. Memorandum on Legal Regulation on the Use of Mobile Positioning Data in the European Union and Estonia, Finland, Germany and France
4) Annex 4. Online Survey Form
5) Annex 5. Expert Interview Guidelines
6) Annex 6. Relevant Patents
7) Annex 7. Aggregated Raw Data Example

8) Annex 8. Table of Responses by Countries and MNOs
9) Annex 9. Analysis of the Survey Results and Interviews with Stakeholders
10) Annex 10. The Initial Responses of the DPA in Finland

1.2. Background of the Report

Report 2 is input to serve as a basis for the subsequent Reports 3 through 4. The following overview shows its position and objective in the overall structure of this feasibility study:

a) Report 1 – Stock-taking contains an up-to-date description of the state of the art in using mobile positioning data in research and applications in tourism statistics and related domains. Report 1 provides a list and descriptions of the experiences of consortium members and other public and private institutions that have been involved in projects where mobile positioning data has been accessed. The report provides references to existing problems and solutions in technology, methodology, regulations and other aspects of accessibility to the data that will serve as important input for Report 2.
b) Report 2 − Feasibility of access will provide an overview of the regulatory, financial, technical and other related topics that cover the aspects of data accessibility using among other things references to the existing project from Report 1. Report 2 provides input for barriers as well as opportunities on this topic.
c) Report 3a − Feasibility of use, methodological issues: This report provides a methodology for the production of tourism statistics by using mobile positioning data. A detailed description of the production process is given. An evaluation of the quality of the described methodology is conducted.
d) Report 3b − Feasibility of use, coherence: Report 3b provides insights to evaluation of coherence between mobile positioning data and statistics from other sources. This report is closely related to Report 3a where methodological differences can produce different results.
e) Report 4 – Opportunities and benefits: this report concentrates on the potential opportunities and benefits of the usage of mobile positioning data for the tourism statistics point of view. In Report 4, the consortium does not collect or research new data and information, but rather integrates the results from previous work packages into a structured and coherent assessment of potential opportunities and benefits.

2. Methodology of the Report

The results for the current report are obtained from the experience and information of consortium members, legal analyses, online survey and interviews with experts on the subject and practical experience of obtaining the data from MNOs (the pilot data within the scope of the current study and experience from projects referenced in Report 1). The report is a combined knowledge set acquired from the mentioned sources divided between specific domains (privacy and regulation-related, financial and business-related and technology-related subjects).

2.1. Online Survey and Expert Interviews

The key objective of the online survey is to learn the state of mind of the potentially involved parties in the chain of value in obtaining and implementing mobile positioning data (mainly in the tourism industry) and to understand specific barriers and solutions in gaining access to the data from the user and provider side.

The survey was carried out between June 2013 and November 2013. Non-probability purposive method and contact networks were used to choose the respondents in this online survey. Requests to participate in the survey were sent to 422 respondents representing experts and stakeholders in Europe (see Figure 1). The survey was open and promoted among potentially interested stakeholders in order to get as much valuable respondents as possible during the project. The consortium partners distributed the survey to mailing lists of thematic networks with size unknown (close to a thousand worldwide).

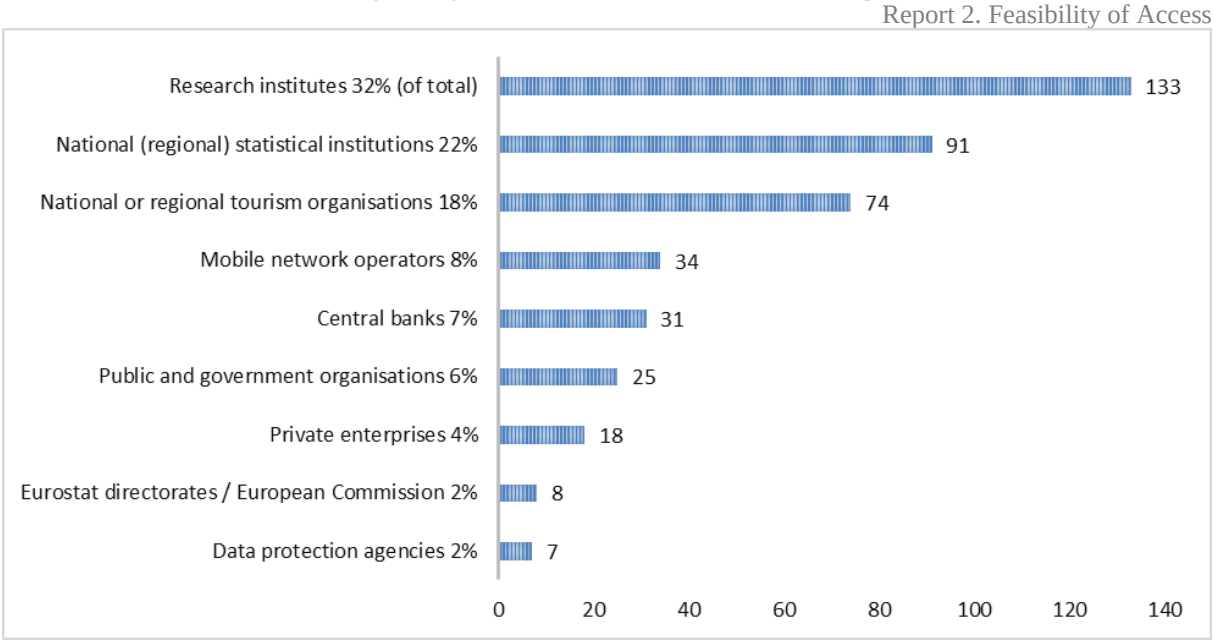

Figure 1 Distribution of the survey sample by type of organisation (n=422, excluding mailing lists).

The survey was divided into 5 parts, which were targeted to the ten main groups of different organisations and stakeholders. Altogether the questionnaire included 73 questions, but not everybody had to answer all of them. MNOs, data protection agencies (DPA) and other stakeholders (national statistical institutes, tourism boards, central banks, private companies etc.) each had separate specific questions. The online survey form is added as Annex 4 to this report.

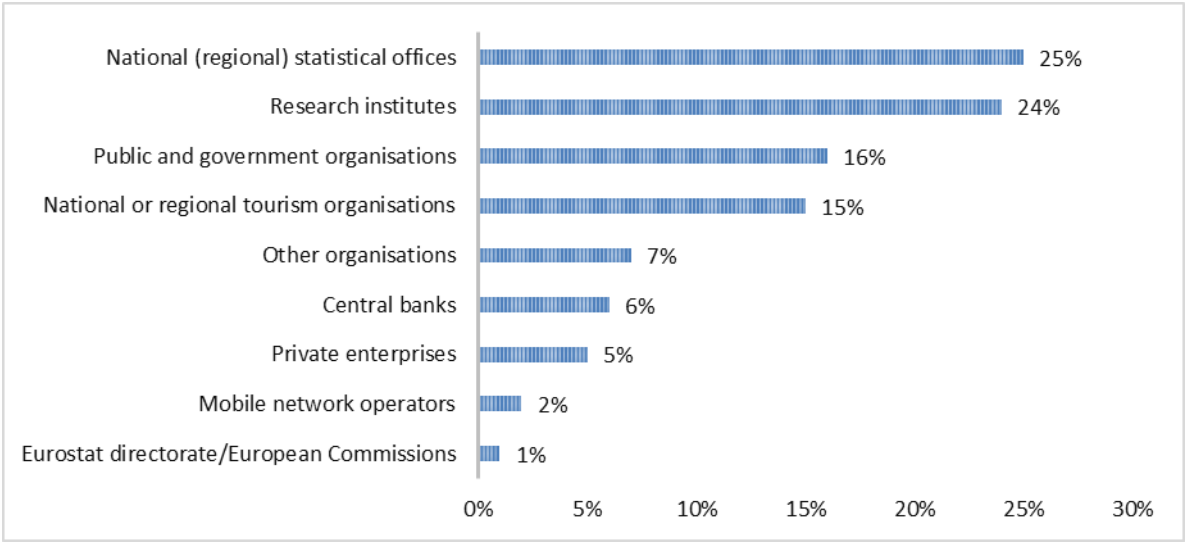

Figure 2. Distribution of the respondents of the online survey by type of organisation (n=118).

The survey garnered 118 responses from 115 organisations. Most responses were given by the representatives of national (or regional) statistical institutes (30, comprising 25% of total) and Research Institutes (29, comprising 24% of total) (see Figure 2). Out of NSIs in

the EU, the survey received responses from 16 out of the total 28. Survey responses came from at least 35 different countries (some respondents' countries were not identifiable). Responses originated from 24 EU Member States (all except Cyprus, Greece, Romania and Slovakia), 10 other European countries and territories (Norway, Iceland, Greenland, Switzerland, Serbia, Montenegro, Bosnia and Herzegovina, Albania, Georgia, Monaco) and 4 non-European countries (Argentina, Colombia, Israel and Japan).

The main objective of the expert interviews was to go deeper than in the online survey and investigate the possibility to use mobile data for the production of statistics in tourism and other areas. The purpose was to get a thorough understanding of topics related to availability of the data and accessibility to it from privacy protection, regulatory, legislative, business, financial and technological points of view.

Most of the interviews were carried out via telephone or Skype in June 2013 – November 2013 in parallel with the online survey. The snowball sampling method was used when searching the experts with deeper knowledge and experiences related to the subject. Some of the experts for the interviews were chosen based on the online survey responses and via personal contacts. A total of 37 interviews were carried out during the project (7 DPAs, 11 MNOs, and users of data – 5 NSIs, 4 research institutes, 1 central bank and 9 public organisations for tourism).

Interviews with MNOs were conducted mostly in parallel with the discussion on access to the data through a pilot project. The main topics emphasised were the regulatory frames and limits that MNOs can act within (in providing access to the data), the essence of the data that can be used in tourism statistics (the format of the data, level of anonymity), technological processes that MNOs require to be implemented in order to provide such data, financial limitations, business secrets and MNOs' internal interests for the use of this data. The table in Annex 8 provides an overview of the number of MNOs in each European country, and the number contacted and discussed with in connection to this study.

MNOs are mostly sympathetic to the idea of using mobile data in tourism statistics – this topic and the value from this methodology are easily understandable. Mostly MNOs expressed concerns regarding legal limitations and obligations to provide the data to the state and other users; public opinion and a possible decline in the number of clients due to bad reputation; technological implementations required to provide the data and its cost; and value for the MNOs if they provide the data.

Other stakeholders involved in the interviews were mostly official tourism authorities, data protection agencies and organisations that have already accessed or are working towards accessing mobile data; therefore, possessing valuable information with regard to the current study. The main objective of the interviews here is to map all the limitations and possibilities concerning the accessibility to the data in their country.

According to the survey results, 89% of the respondents are using mobile positioning data for tourism statistics, indicating that the survey is delivered to appropriate experts and stakeholders. Despite the high awareness (86%) of the possibilities provided by mobile positioning data, mere 14% of all respondents actually use it. Among those, some use phone data using special applications installed on devices. The users include mostly research institutes and to a lesser amount private enterprises. Only some NSIs or other governmental organisations (municipalities) are using such data in tourism. Though the interest towards using the data is high, the responses indicate that the access to the data is very limited due to regulatory barriers (see Figure 3).

Based on the survey and interviews, mobile positioning data is used in some form in Spain, the Netherlands, Czech Republic and Italy. However, only aggregated form of raw data is being analysed currently. First steps have also been taken in many countries to start using mobile positioning data in tourism statistics, e.g. in Ireland, Slovenia, Belgium, Austria and Greenland.

Survey results and key findings from interviews are presented within the respective sections of this and other reports of the current study. The survey questions and the guidelines of the interviews are presented in Annexes 4 and 5 of the current report. An overview of MNO contacts and NSI interest in the use of mobile positioning data can be found in Annex 8. Analysis of the results of the survey and interviews is presented separately in Annex 9 of the current report.

2.2. Legal Analysis

Legal aspects of accessibility to the data are one of the main questions to be addressed in this study. This is also indicated in the responses from the survey and interviews. Legal experts from four EU countries are included in this study to analyse the limitations and opportunities of EU directives and national-level legislation on the subject of using mobile positioning data in general and in tourism statistics. The legal analyses for this report were carried out by Attorneys at Law Borenius (Estonia), Beiten Burkhardt (Germany), Astine

Avocats (France) and internal legal experts from Statistics Finland, Statistics Estonia and National Institute of Statistics and Economic Studies France (INSEE).

EU directives act as an umbrella for national legislation and local governments mostly implement the directives with local nuances. Therefore, it is assumed to be possible to describe the overall effect of the EU directives and analyse the practical variations of the legislation in four countries (partner countries of this consortium – Estonia, Finland, France and Germany). Analysis of practical implementation in every EU country is not in the scope of this study; therefore, the current analysis will provide a description of the variability of the implementations of the directives and not the full scope of the applicability in EU countries.

The legal analysis of the current report (see Section 3) concentrates on the EU directives that are relevant to the subject and the national variability in the application of such directives based on the example of four countries.

3. Privacy and Regulation-related Opportunities and Barriers

The analysis on the EU and Member States legislation is based on the memorandum provided to the consortium by the team of external legal experts and is slightly adapted to current report. The memorandum is structured by first giving a list of the relevant EU legal acts and the subject matter governed by them. This is followed by sections mapping the relevant laws of each involved sample jurisdiction providing the general framework as well as a short overview of the applicable rules. Thirdly, a chapter discussing in more detail the main issues of concern that arose during conducting the analysis is provided. The memorandum is summed up by drawing conclusions based on the jurisdictions and implementation practice in sample countries as well as giving recommendations on what additional measures should be taken in order to clarify the applicable legal framework and eliminate risks highlighted.

Privacy concerns and legislation are the main barriers to obtaining the data from MNOs according to the responses from the survey and interviews (see Figure 3). The legal framework governing the use of location data within the European Union is rather complex despite having the two directives of the European Parliament and of the Council directly governing the topic under discussion in force for a while. The complexity of ever developing technological solutions used and business models applied leads to difficulties in the categorisation of actual services and players in commercial chains according to the directives

and relevant local laws. At the same time the lawfulness of data processing largely depends on such categorisation.

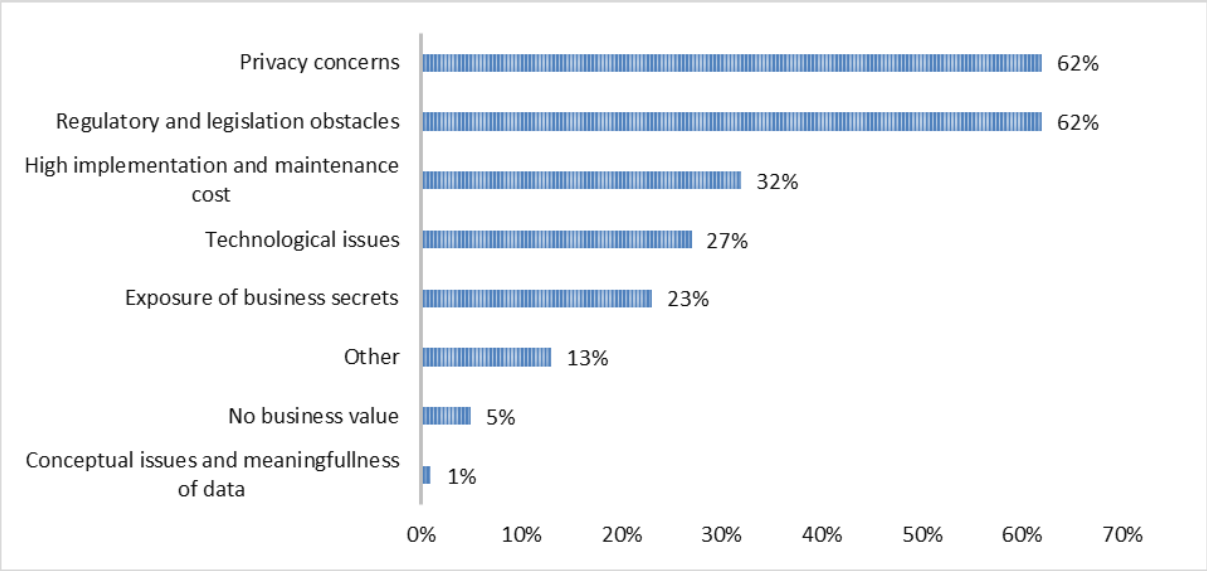

Figure 3. Main obstacles in getting data from MNOs according to responses from the survey (n=116).

For official statistics, a national statistics act is a legal instrument for accessing the data from enterprises (MNOs); however, there is no certainty of its practical implementation regarding the specificities of the data that can be provided by the MNOs. There are also contradictions in responses from the survey concerning the power of such instrument – some suggest the current legislation is powerful enough to 'force' the MNOs to provide all the necessary data, and privacy protection is of no concern; some suggest that although the act is powerful enough, there should be acceptance from a national privacy protection body and specification of the data in use before implementing the act; and some suggest that the current legislation does not provide sufficient means and requires new legislation initiatives. Others question whether data protection means that every person has the right to decide whether he or she wants to be included in Statistics or not. Yet others say the question who can have access to mobile positioning data has to be treated as a social issue by ethical committees rather than by legal officers, the last are more formally oriented and prefer to be conservative. Some respondents from the tourism statistics community propose the creation of unified legal and practical methods of accessing and using the data that would be in accordance with EU and national legislation as well as the principle methodological approach everywhere. One respondent mentions: 'It is important that the Statistical Community work to establish a right of access to these sources in principle for Official Statistics purposes.'

The current chapter concentrates on the aspects of accessibility of the data from legal and privacy protection aspects. An overview on the relevant EU legislation, the legislation of

the four sample EU member countries as well as analysis of the major issues of concern is given below.

3.1. European Union Legislation

The links to the regulatory documents are provided in Annex 2 to the current report.

3.1.1. Effective Instruments

The currently effective legal acts on the EU level directly relevant to the topic at hand are the following.

3.1.1.1. Directive 1995/46/EC

Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (hereinafter the DPD).

Also referred to as the 'Data Protection Directive', the DPD is the fundamental instrument for the protection of personal data in the EU setting out the concept of personal data as well as the general principles and rules on the processing of such data. The DPD applies in every case where personal data is being processed as a result of the processing of location data.

3.1.1.2. Directive 2002/58/EC

Directive 2002/58/EC (as amended by 2009/136/EC) of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (hereinafter the EPD).

Also referred to as the 'E-privacy Directive' the directive deals with privacy and personal data protection matters in the electronic communications sector. Consequently, the EPD only applies to the processing of the data processed by the telecom operators supplementing and specifying the DPD in aspects specific to the telecommunications sector.

3.1.1.3. Directive 2006/24/EC

Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC (hereinafter the DRD).

Also referred to as the 'Data Retention Directive', the directive requires electronic communications operators to store certain categories of their customers' data for a period of six months to two years to make them available, upon request, to law enforcement authorities for the purposes of the investigation, detection and prosecution of serious crime.

3.1.1.4. Article 29 Data Protection Working Party Opinion

Article 29 Data Protection Working Party Opinion 13/2011 of 16 May 2011 on Geolocation services on smart mobile devices (hereinafter the WP29 Opinion).

The Article 29 Data Protection Working Party (hereinafter the Working Party) is a body set up under Article 29 of the DPD consisting of a representative of the supervisory authorities of each Member State, the European Data Protection Supervisor, and of a representative of the European Commission. The Working Party has advisory status and acts independently. Among other, the Working Party examines any question covering the application of the national measures adopted under the DPD in order to contribute to the uniform application of such measures as well as makes recommendations, on its own initiative, on all matters relating to the protection of persons with regard to the processing of personal data in the Community. The opinions of the Working Party are of advisory nature. However, since the Working Party is composed of the representatives of the supervisory authorities on both the EU and Member States level, the Working Party's opinions should be regarded as reflecting the current position on data protection issues of the supervisory authorities on applying the effective legal acts. Therefore, it is recommended to take note of the Working Party's opinions including the WP29 Geolocation Opinion. As to the use of the geolocation data, the WP29 Geolocation Opinion is the main document addressing respective issue in detail in light of the relevant EU directives in force.

In addition to the above listed instruments there are others that govern certain issues that may be of relevance in terms of the use of mobile positioning data. References to such legal acts or other documents are made in the report where appropriate.

3.1.2. Proposed Instruments

3.1.2.1. General Data Protection Regulation

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (hereinafter the Draft Regulation).

In addition to the legislation in force it should be taken into account that the European Commission has initiated the EU data protection reform by way of disclosing the said proposal for a new regulation that is also referred to as the General Data Protection Regulation.

Although the Draft Regulation contains, as does the DPD, a special article on the processing of personal data for statistics purposes, it must be noted that the Draft Regulation in its current working version does not add too much clarity as to the terms governing the processing of location data.

Although the language of the Draft Regulation makes use of the term 'location data', as opposed to the DPD, it is used only to indicate that location data may but need not necessarily be considered personal data. Accordingly, Recital 24) stipulates the following: 'When using online services, individuals may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses or cookie identifiers. This may leave traces which, when combined with unique identifiers and other information received by the servers, may be used to create profiles of the individuals and identify them. It follows that identification numbers, location data, online identifiers or other specific factors as such should not be considered as personal data if they do not identify an individual or make an individual identifiable'. Thus, it can be concluded that in the view of the European Commission the location data may but need not be personal data. The categorisation depends on specific circumstances of each case the data is processed.

Moreover, Art 4 (1) of the Draft Regulation uses location data as one type of data via which a data subject can be identified setting out the following: ''personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.'

The processing of personal data for official statistics purposes is mainly governed by Art 83 of the Draft Regulation. Pursuant to Art 83 the personal data may be processed for historical, statistical or scientific research purposes only if:

(a) these purposes cannot be otherwise fulfilled by processing data which does not permit or not any longer permits the identification of the data subject;

(b) data enabling the attribution of information to an identified or identifiable data subject is kept separately from other information as long as these purposes can be fulfilled in this manner.

It is to be considered that the bodies conducting statistical research may publish or otherwise publicly disclose personal data only if:

(a) the data subject has given consent;
(b) the publication of personal data is necessary to present research findings or to facilitate research insofar as the interests or the fundamental rights or freedoms of the data subject do not override these interests; or
(c) the data subject has made the data public.

The Draft Regulation further stipulates that the European Commission shall be empowered to adopt delegated acts for the purpose of further specifying the criteria and requirements for the processing of personal data for the statistics purposes as well as any necessary limitations on the rights of information to and access by the data subject and detailing the conditions and safeguards for the rights of the data subject under these circumstances.

Therefore, although it can be presumed that by Art 83 the legislator indicates that it considers it essential to enable the processing of personal data for statistics purposes on special, and somewhat less strict conditions than those applying to the processing of the same for private purposes, the specific criteria and requirements are still to be set forth by the European Commission by adopting the delegated acts.

The European Parliament's Committee on Civil Liberties, Justice and Home Affairs (LIBE) voted on the Draft Regulation on 21 October 2013. The European Commission currently aims to update the current EU data protection legislation before the next European elections in May 2014 and the regulation is planned to take effect in 2016 after a transition period of two years. However, this entails reaching a trilateral agreement on the final version by the European Parliament, the Council and the Commission as well as the final vote by the European Parliament which is why it is currently unclear when exactly the Draft Regulation is going to be adopted.

3.1.3. General Overview of Legal Framework

3.1.3.1. Main Concepts of Personal Data Protection

3.1.3.1.1. Personal Data

Pursuant to the DPD personal data is any information relating to an identified or identifiable natural person (a data subject) whereas an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

Consequently, the DPD applies to processing of any information that relates to an individual that is identifiable by way of such information or data that is being processed.

In terms of the use of the mobile positioning data it is therefore first essential to identify if the data used qualifies as personal data or not. The DPD (among other, the obligation to obtain a data subject's consent for any type of data processing) does not apply in case the mobile positioning data collected and otherwise processed does not qualify as personal data.

The current study addresses the processing of anonymous and aggregated data. In order to preclude the applicability of the DPD one must determine that the mobile positioning data collected and otherwise processed is at all times throughout the process anonymous and that it cannot at any point be tracked down to an identifiable data subject (see Section 3.3 for further analysis).

In each case where the data collected qualifies as personal data, the provisions of the DPD must be adhered to. These include, among other, the obligation to make sure that data is processed on one of the lawful basis set out in Section 3.1.3.1.4.

3.1.3.1.2. Processing of Personal Data

Under the DPD processing of personal data is to be understood as any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.

Resulting from the above, if the mobile positioning data qualify as personal data at any given point of time, any operations relating to such data (incl. collection, storing, using, etc.)

constitutes the processing of personal data and can only be done in strict compliance with the DPD.

3.1.3.1.3. Data Controller and Data Processor

Under the DPD a data controller is to be understood as the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data.

A natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller is considered a data processor.

Making the distinction between a data controller and a data processor is essential since the data controller must procure that the processing of personal data is conducted according to the law. The data controller must also procure that the data processor that processes personal data on behalf of the controller does so in compliance with the law.

In light of the study at hand, it should be identified which parties involved in the process of positioning data collection and use are considered data controllers and which of them data processors. Based on such identification of function the rights and obligations of each stakeholder can be determined.

3.1.3.1.4. Lawful Bases of Processing Personal Data, Data Subject's Consent

The general principle applying to personal data processing in the EU is that the processing is allowed on a data subject's due consent or in case a statutory basis for using personal data without a data subject's consent exists. Such two conceptual bases have been set out in more detail in Art 7 of the DPD under which personal data may be processed only if:

(a) the data subject has unambiguously given his or her consent; or
(b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or
(c) processing is necessary for compliance with a legal obligation to which the controller is subject; or
(d) processing is necessary in order to protect the vital interests of the data subject; or
(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data is disclosed; or

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data is disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1) of the DPD.

Therefore, unless specific statutory bases for processing exist, a data subject's consent should be obtained for each type of data processing in case the mobile positioning data qualifies as personal data.

The specific statutory bases relevant in the context of processing personal data by the state statistics authorities is that set out in Section 3.1.3.1.4 (e) – processing of personal data is necessary for the performance of a task carried out in the exercise of official authority vested in the data controller or in a third party to whom the data is disclosed.

Therefore, as a rule, the state statistics authorities as the data controllers may process personal data for the performance of a task in the exercise of official authority vested in them. However, the local laws are somewhat ambiguous concerning the applicability of such ground (refer to Section 3.3.3 for details).

3.1.3.1.5. 'Ownership' of Data

From the legal perspective the ownership of personal data is to be understood as a person's right to store, use or otherwise process the personal data relating to a data subject. Such right originally lies with the data subject who may determine the terms and conditions of his or her personal data by others. Such determination is made by way of a data subject's consent for processing his or her personal data. Additionally, personal data may be used according to the consent given by the data subject or under the law if the latter provides such specific basis.

When determining the rights to data other than personal data it must be taken into account that mobile positioning data generally constitutes facts and other factual data. Facts and information as such are not protected by copyright. It is likely, however, that the owner of the database containing such facts or information enjoys the protection of a database owner. Since the use of the database owner's rights is a matter to be contractually handled between the owner of a database (e.g. an MNO) and the user of a database (e.g. NSI, data broker, Eurostat, etc.) and it does not involve consideration of statutory rights of a data subject, the database licensing rights will not be discussed here.

3.1.3.1.6. Location Data and Traffic Data and Processing thereof

According to Art 2 (c) of the EPD location data means any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service. Recital (14) of the EDP stipulates that location data may refer to the latitude, longitude and altitude of the subscriber's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded.

Pursuant to Art 2 (b) of the EPD traffic data is defined as any data processed for the purpose of the conveyance of a communication on an electronic communications network or for the billing thereof.

Under Art 9 (1) of the EDP location data other than traffic data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service (see Section 3.1.3.1.7). The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. Users or subscribers shall be given the possibility to withdraw their consent for the processing of location data other than traffic data at any time.

Where consent of the users or subscribers has been obtained for the processing of location data other than traffic data, the user or subscriber must continue to have the possibility, by simple means and free of charge, of temporarily refusing the processing of such data for each connection to the network or for each transmission of a communication (EPD, Art 9 (2)).

Processing of location data other than traffic data in accordance with Art 9 (1) and (2) must be restricted to persons acting under the authority of the provider of the public communications network or publicly available communications service or of the third party providing the value added service, and must be restricted to what is necessary for the purposes of providing the value added service.

Traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must

be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication except as follows:

(a) traffic data necessary for the purposes of subscriber billing and interconnection payments may be processed. Such processing is permissible only up to the end of the period during which the bill may lawfully be challenged or payment pursued;
(b) certain traffic data may be used for the purpose of marketing electronic communications services or for the provision of value added services if the subscriber or user to whom the data relates has given his/her consent.

3.1.3.1.7. Value Added Services

Pursuant to Art 2 (g) of the EPD a value added service means any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof.

3.1.3.1.8. Mobile Operators' Obligation to Retain Customer Data

Pursuant to Art 1 (2) of the DRD the latter applies to traffic and location data on both legal entities and natural persons and to the related data necessary to identify the subscriber or registered user.

Under Art 4 of the DRD the Member States must adopt measures to ensure that data retained in accordance with the DRD is provided only to the competent national authorities in specific cases and in accordance with national law. The procedures to be followed and the conditions to be fulfilled in order to gain access to retained data in accordance with necessity and proportionality requirements shall be defined by each Member State in its national law, subject to the relevant provisions of European Union law or public international law, and in particular the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) as interpreted by the European Court of Human Rights. Therefore, the extent and terms on which the data retained by the MNOs may be disclosed to third persons, including the competent authorities, must be stipulated by each local Member State law.

According to Art 5 of the DRD the following categories of data are to be retained concerning mobile telephony:

(a) data necessary to trace and identify the source of a communication
    (i) the calling telephone number;
    (ii) the name and address of the subscriber or registered user;
(b) data necessary to identify the destination of a communication:

(i) the number(s) dialled (the telephone number(s) called), and, in cases involving supplementary services such as call forwarding or call transfer, the number or numbers to which the call is routed;

(ii) the name(s) and address(es) of the subscriber(s) or registered user(s);

(c) data necessary to identify the date, time and duration of a communication:

(i) the date and time of the start and end of the communication;

(d) data necessary to identify the type of communication:

(i) the telephone service used;

(e) data necessary to identify users' communication equipment or what purports to be their equipment:

(i) the calling and called telephone numbers;

(ii) the International Mobile Subscriber Identity (IMSI) of the calling party;

(iii) the International Mobile Equipment Identity (IMEI) of the calling party;

(iv) the IMSI of the called party;

(v) the IMEI of the called party;

(vi) in the case of pre-paid anonymous services, the date and time of the initial activation of the service and the location label (*cell_id*) from which the service was activated;

(f) data necessary to identify the location of mobile communication equipment:

(i) the location label (*cell_id*) at the start of the communication;

(ii) data identifying the geographic location of cells by reference to their location labels (*cell_id*) during the period for which communications data is retained.

The listed data must be retained by the communications services providers for the period of six months to two years as from the date of communication. The exact terms of data retention within this range are to be enacted by Member States in their local laws.

It is important to note that the data retained according to the DRD is subject to all requirements of personal data processing and the retention thereof does not release the data controller from adhering thereto, i.e. the retained data cannot be processed (in any other way that the retaining thereof) without the data subject's due consent or under express statutory basis.

3.2. National Legislations

The below chapter will give an overview of the state of legislation in four countries – Estonia, Finland, France and Germany – that should provide an example. A table in Annex 8

summarises the state of transposition of the Data Retention Directive in European countries and the variation in the ascribed minimum period for retaining mobile network data country by country.

3.2.1. Estonia

The following Estonian legal acts are applicable to the collection and use of mobile positioning data:

3.2.1.1. Personal Data Protection Act (hereinafter the PDPA)

The PDPA is the instrument by which the DPD has been transposed into Estonian law. The PDPA sets out the main terms and principles of processing personal data in line with the DPD.

3.2.1.2. Electronic Communications Act (hereinafter the ECA)

The ECA is the instrument by which the EPD and the DRD have been transposed into Estonian law. The ECA sets out general rules of processing subscribers' location data by mobile operators.

Pursuant to Art 105 (1) of the ECA a communications undertaking has the right to process subscribers' location data, the processing of which is not provided for in Art 104 (data necessary for billing the subscriber) or Art 111^1 (data subject to data retention obligation) of the ECA, only if such data is rendered anonymous prior to processing.

A communications undertaking (an MNO) may also process, with the consent of the subscriber, the data provided for in the previous paragraph to provide other services in the process of using the communications services to an extent and during the term necessary for processing and without rendering the data anonymous.

3.2.1.3. Data Retention Obligation

Under Art 111^1 (1) of the ECA any communications undertaking is required to retain the data that is necessary for the performance of the following acts:

1) tracing and identification of the source of communication;
2) identification of the destination of communication;
3) identification of the date, time and duration of communication;
4) identification of the type of communications service;

5) identification of the terminal equipment or presumable terminal equipment of a user of communications services;
6) determining of the location of the terminal equipment.

More specifically, the providers of mobile telephone services and mobile telephone network services are required to preserve the following data:

1) the number of the caller and the subscriber's name and address;
2) the number of the recipient and the subscriber's name and address;
3) in the cases involving supplementary services, including call forwarding or call transfer, the number dialled and the subscriber's name and address;
4) the date and time of the beginning and end of the call;
5) the telephone or mobile telephone service used;
6) the international mobile subscriber identity (IMSI) of the caller and the recipient;
7) the international mobile equipment identity (IMEI) of the caller and the recipient;
8) the *cell_id* at the time of setting up the call;
9) the data identifying the geographic location of the cell by reference to its *cell_id* during the period for which data is preserved;
10) in the case of anonymous pre-paid mobile telephone services, the date and time of initial activation of the service and the *cell_id* from which the service was activated.

The data must be retained for one year from the date of the communication if such data is generated or processed in the process of provision of communications services.

The data controller must ensure that the processing of the retained data complies with general personal data processing rules. Therefore an MNO may process the data that the MNO must retain under the data retention obligation in compliance with general personal data processing rules. Hence, the retained data may be processed if the data subject has given consent or if there is a specific statutory basis for such processing.

3.2.1.4. Official Statistics Act (hereinafter the OSA)

The OSA governs the use of personal data in production of official statistics. Under Section 31 of the OSA a producer of official statistics has the right to use personal data on the bases of and pursuant to the procedure provided for in the PDPA in the production of official statistics. A producer of official statistics is not required to inform data subjects of the use of their personal data in producing official statistics. Please refer to Section 3.3.3 on the

provisions of the PDPA relating to the processing of the personal data for the purposes of official statistics and relation thereof to Section 31 of the OSA.

3.2.2. Finland

The following Finnish legal acts are applicable to the collection and use of mobile positioning data.

3.2.2.1. Personal Data Act (hereinafter the HTL)

The HTL is the instrument transposing the DPD into the Finnish legislation. It contains the terms and conditions and general principles of processing personal data in Finland.

The definitions used for personal data and the processing of personal data in the HTL are identical to the DPD. Art 8 (1) and (4) of the HTL set out general prerequisites to processing of personal data. The relevant prerequisites are:

(1) the data subject has unambiguously consented to the same;
(4) processing is based on the provisions of law or it is necessary for compliance with a task or obligation to which the controller is bound by virtue of law or an order issued on the basis of law;

3.2.2.2. Act on the Protection of Privacy in Electronic Communications (hereinafter the SVTL)

The SVTL is the instrument by which the DRD is transposed into Finnish legislation. An amendment to the act was made in 2008 making it compliant with the DRD.

3.2.2.3. The Statistics Act (hereinafter the TL)

The TL lays down provisions on the procedures and principles concerning the collection of data and the designing and production of statistics that shall be applied by state authorities in their statistics compilation. The collection, release, protection and other processing of data during the compilation of statistics must be subject to the provisions of the Act on the Openness of Government Activities (621/1999) and of the Personal Data Act (523/1999), unless provided otherwise by the law.

On data collection Art 5 of the TL stipulates that the data must be collected and stored without identification data whenever permitted by the statistics to be produced. Identification data may only be collected and stored where it is necessary for data linking or when otherwise

deemed necessary for the production of reliable and comparable statistics depicting features in the development of social conditions.

Art 10 of the TL stipulates that when data collected for statistical purposes is being combined, stored, destroyed or otherwise processed it shall be ensured that no person's protection of private life or personal data, or business or professional secret shall be endangered.

Concerning the right of Statistics Finland to collect data from enterprises, Art 14 of the TL provides that they are obliged to provide Statistics Finland with data on the type, location, ownership, finances and products of their activities, as well as with data on the staff and other resources required in their activities. However, the obligation of enterprises does not extend to providing data on the users of their products or services. Therefore based on the current TL enterprises cannot be obliged to provide data on their client registers such as client positioning data collected by MNOs.

3.2.2.4. Communications Market Act (hereinafter the VML)

The objective of the VML is to promote the provision and use of services within communications networks and to ensure that communications networks and communications services are available under reasonable conditions to all telecommunications operators and users throughout the country.

The VML act established the legal framework for the operations of telecommunications service providers in Finland. It does not directly deal with issues regarding use of personal data or positioning data.

3.2.3. France

The following French legal acts are applicable to the collection and use of location data:

3.2.3.1. Information Technology, Data Files and Civil Liberties Act (hereinafter the ITLA)

The ITLA transposed the DPD into French law and sets out the main terms and conditions concerning the protection of personal data in France. The French data protection regulator (hereinafter the CNIL) also regularly issues recommendations and guidelines. Such recommendations and guidelines are not mandatory rules but should generally be followed unless this is justified by a legitimate ground and in compliance with the ITLA.

According to Section 2 (2) of the ITLA, personal data means any information relating to a natural person who is or can be identified, directly or indirectly, by reference to an identification number or to one or more factors specific to him. In order to determine whether a person is identifiable, all the means that the data controller or any other person uses or may have access to should be taken into consideration.

Location data is generally considered by the CNIL as personal data, insofar as it can be related to the subscriber of an electronic communications service.

Anonymous data is not defined as such by the ITLA but the CNIL issued guidelines on anonymisation in its 2010 Guide on Security of Personal Data.

In order not to be subject to the ITLA, processing of personal data must be subject to an irreversible anonymisation, which consists in removing any identifying character from a set of data. This means that all directly and indirectly identifying information is removed and that it is impossible to re-identify the persons. The CNIL recommends to:

- be very careful insofar as re-identification can take place from partial information;
- proceed as follows to anonymise personal data:
    - generate a secret that is long enough and difficult to memorise;
    - apply a 'one-way' function to the data: an algorithm suitable for such an operation is a keyed hash function such as the HMAC algorithm based on SHA-1.

The CNIL considers that anonymisation mechanisms that have not been validated by experts should not be used. In particular a good anonymisation algorithm must:

- be irreversible;
- present a very weak collision rate: two different sets of data should not lead to the same result;
- present a great dispersion: two quasi-similar sets of data must have very different results;
- use a secret key.

In some cases the CNIL considers that a double reversible anonymisation is advisable, i.e. the application of a second anonymisation on the result of a first anonymisation, both anonymisations using different secrets, held by separate organisations. The FOIN algorithm

(Fonction d'Occultation des Informations Nominatives, Personal Information Hiding Function) is an example of algorithm using double anonymisation.

The principles according to which personal data can be lawfully processed are the same as the principles set out by the DPD.

In particular, Art 6 (2) of the ITLA provides that the data shall be obtained for specified, explicit and legitimate purposes, and shall not subsequently be processed in a manner that is incompatible with those purposes. However, further data processing for statistical, scientific and historical purposes shall be considered compatible with the initial purposes of the data collection, if it is carried out in conformity with the principles and procedures provided for in Chapter II of the ITLA (conditions for a lawful data processing which are similar to the conditions set out in the DPD), in Chapter IV (formalities, e.g. notifications, prior to commencing data processing) and in Art 1 of Chapter V (obligations incumbent upon data controllers and rights of data subjects, mainly relating to the information of the data subjects, the security of the data and the retention duration) and if it is not used to take decisions with respect to the data subjects.

The consent of the data subject is necessary for the processing of his or her data, under the same conditions as those set out in the DPD.

In this respect, under Art 6 of the ITLA, consent of the data subject is not necessary for data processing relating to compliance with any legal obligation to which the data controller is subject or the performance of a public service mission entrusted to the data controller or data recipient.

If the data is not anonymised when transferred to the statistics authorities and the processing is therefore subject to the ITLA, a standard notification should be filed with the CNIL by the data controller describing, notably, the purpose of the processing, the categories of data subject to the processing, together with the recipients and the retention duration, the security and confidentiality measures in place and how the data subjects are informed of their rights. The data processing cannot commence before receiving the CNIL's receipt of such notification. In the case contemplated in this study, the statistic authorities could be considered as data controller.

The notification does not exempt the data controller from respecting its other obligations, in particular from obtaining the data subjects' consent when necessary.

Prior authorisations are necessary for some data processing (in particular in case of sensitive data processing, interconnection of databases having different purposes, transfers

outside of the European Union or processing used to take decisions regarding the data subjects). This point should be further reviewed, but if the intent is to only transfer identification data and location data for statistical purposes and the data remains within the European Union, no specific authorisation should be necessary.

Art 32.III of the ITLA provides that whenever the data has not been obtained from the data subject, the data controller or its representative must, at the time of recording the personal data or, if disclosure to a third party is planned, no later than the time when the data is first disclosed, provide the data subject with the following information:

- the identity of the data controller and of his representative, if any;
- the purposes of the processing for which the data is intended;
- whether replies to the questions are compulsory or optional;
- the possible consequences for him of the absence of a reply;
- the recipients or categories of recipients of the data;
- the rights of individuals in relation to the processing of data (i.e. right of access, rectification and opposition);
- when applicable, the intended transfer of personal data to state that is not a Member State of the EU.

When the personal data has initially been obtained for another purpose, the above provisions shall not apply to processing necessary for the storage of this data for historical, statistical and scientific purposes, under the conditions provided for in Book II of the Heritage Code or for the re-use of this data for statistical purposes under the conditions provided for in Art 7 bis of Act No. 51-711 of 7 June 1951 on obligations, co-ordination and confidentiality as regards statistics. Book II of the Heritage Code relates to archiving of data and Art 7 bis of Act No. 51-711 relates to the assignment to the French state statistic authorities of personal data collected by a public or private entity managing a public service.

In addition, the above obligations of information do not apply whenever the data subject has already been informed or whenever informing the data subject proves impossible or would involve disproportionate efforts compared with the interest of the procedure.

Finally, Art 32.IV of the ITLA provides that if the personal data obtained is, within a short period of time, to form part of an anonymisation procedure that was recognised beforehand by the CNIL as complying with the provisions of the ITLA, the information delivered by the data controller to the data subject may be limited to that mentioned in subsections above.

Art 38.II of the ITLA also provides that the right of access of the data subjects is not applicable when the personal data is stored in a form that clearly excludes all risk of violating the privacy of the data subject and for a period that does not exceed that necessary for the sole purpose of creating statistics, or for scientific or historical research. Such exemptions by the data controller must be mentioned in the application for authorisation or in the notification addressed to the CNIL.

According to Art 36 of the ITLA, personal data may be stored beyond the period necessary for the purpose of the processing only for historical, statistical and scientific purposes. The choice of the data that is stored must be made, in such a case, in accordance with the legal provisions relating to public data (Art L212-4 of the Heritage Code).

3.2.3.2. Postal and Electronic Communications Code (hereinafter the PECC)

The PECC sets out the main terms and conditions applicable to electronic communications in France, including the processing of location data and other data collected by MNOs.

The articles of the PECC relating to personal data processing apply to the providers of electronic communications services and to networks which are in charge of data and identification collection tools.

In accordance with Art L34-1.II of the PECC, the MNOs must erase or make anonymous any data relating to traffic, i.e. any information made available through electronic communications means that may be recorded by the MNO during the electronic communications transmitted though the MNO. The exceptions to such obligations are:

- data retention obligation, as described below,
- for invoicing purposes, for certain categories of data,
- for the purpose of providing value added services, as described below,
- for the location data, under certain conditions, as described below, and
- for some data, for the purpose of ensuring the security of their network.

MNOs are required to retain data (including location data), pursuant to Art L34-1.III of the PECC, but only for the purpose of searching, observing and prosecuting criminal offenses or a breach of copyrights, and such data can only be communicated to a judicial authority or the authority in charge of monitoring breaches of copyright. Such data must be kept for only one year.

Art L34-1.IV provides that MNOs can process the data relating to traffic for the purpose of commercialising their own electronic communications services or to provide added value service, only with the consent of the data subject and for a limited duration. Such duration cannot exceed the period necessary for the supply or commercialisation of the services.

Art 34-1.V of the PECC also provides that location data cannot be used during communication for any purpose other than the routing of the same, nor be kept nor stored after the end of the communication without the consent of the subscriber, duly informed of the categories of data in question, the duration of the processing, its purposes and the fact that such data will or will not be transferred to third party service providers. The subscriber must be able to withdraw his consent at any time without cost, except for the cost linked to the transmission of the withdrawal. The subscriber must also be able to suspend his consent by a simple and free of charge mean, except for the cost linked to the transmission of such suspension. The prior consent of the data subject is consequently necessary for personal data to be transferred from the MNOs to the statistical authorities.

Art 34-1.VI of the PECC further provides that the data kept and processed under the above conditions (both under Art 34-1.IV and Art 34-1.V) can only exclusively concern the identification of the persons using the services provided by the MNOs, the technical characteristics of the subscribers' communications and the location data.

They can in no case relate to the content of the exchanged correspondences or the information consulted by the subscribers, in any form whatsoever, during such communications.

The processing of the data must also comply with the ITLA and the MNOs must take all appropriate measures to prevent use of the data for other purposes than the above.

3.2.4. Germany

The following German legal acts are applicable to the collection and use of mobile positioning data:

3.2.4.1. Federal Data Protection Act (hereinafter the BDSG)

The BDSG sets out the main terms and conditions as well as the general principles of processing personal data in Germany. The BDSG constitutes the instrument of transposition of the DPD into German law.

According to Art 4 (1) of the BDSG the collection, processing and the use of personal data shall be lawful only if permitted or ordered by the BDSG or any other law, or if the data subject has provided its consent.

Personal data means any information concerning the personal or material circumstances of an identified or identifiable natural person (a data subject). If information is anonymised (with no possibility to re-identify the respective person), the limitations and requirements of data protection, in particular those arising under the BDSG, are not applicable.

Under Art 3 (6) of the BDSG 'rendering anonymous' is the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense and effort. There is no specific case law available providing guidelines as to what is deemed a disproportionate effort. However, the common understanding is that a conservative and strict approach is to be taken upon construing the said concept with even the slightest possibility to re-identify the natural person prevents the qualification of data as anonymous data.

Where other federal laws apply to the processing of personal data, such as the TKG, they take precedence over the provisions of the BDSG.

3.2.4.2. Telecommunications Act (hereinafter the TKG)

The TKG sets out the main terms and conditions, as well as the general principles of the rights and obligations regarding the provision of telecommunication services in Germany. It also contains special provisions for processing personal data related to telecommunication. The TKG transposes the EPD into German law for telecommunication related issues.

The TKG also used to be the instrument by which the DRD was transposed into German law. However, due to a decision of the Federal Constitutional Court of Germany of March 2, 2010, the clauses which determined the conditions for the obligation of electronic communication operators to store certain categories of personal data to make them available, upon request, to law enforcement authorities for the purposes of the investigation, detection and prosecution of serious crimes, have been considered as a violation of the German Constitution and in consequence revoked and declared invalid.

Despite several legislative initiatives and a pending action of the European Commission before the European Court of Justice, Germany has still not transposed the DRD into German law.

Art 96 (1) No. 1 of the TKG stipulates that the service provider itself may, without the data subject's consent, collect and use the following traffic data to the extent required for the purposes of establishing, framing the contents of, modifying or terminating a contract for telecommunications services:

1) number or other identification of the telecommunication connections involved or of the terminal equipment, personal authorisation codes, card number (if customer cards are used) and additionally, when mobile devices are used, the location data;
2) start and end of the connection, indicated by date and time and, if relevant for charging purposes, the volume of data transmitted;
3) telecommunication service used by the user;
4) termination points of fixed connections, beginning and end of their use, indicated by date and time and, if relevant for charging purposes, the volume of data transmitted;
5) any other traffic data required for setup and maintenance of the telecommunications connection and for billing purposes.

Without the traffic data stored by the telecommunications service provider personal data may be used after the termination of a connection only when required to set up a further connection or for the purposes of:

1) invoicing of the services (Art 97 TKG);
2) provision of an itemised bill (Art 99 TKG);
3) identification and elimination of faults and malfunctions, identification of misuse (Art 100 TKG);
4) call tracing in case of telephone stalking (Art 101 TKG).

Otherwise traffic data is to be erased by the service provider without undue delay after the termination of the connection.

Special limitations regarding the use of location data which is not traffic data are contained in Art 98 of the TKG. According thereto location data relating to users of public telecommunications networks or publicly available telecommunications services may only be

processed when made anonymous or with the customer's consent to the extent and for the duration necessary for the provision of value added services. In this case the service provider has to send a text message to the respective mobile device each time the respective device is being located if the location data is being displayed on a device other than the located mobile device. If location data is transmitted to a third party (other than the provider of the value-added service) the customer's consent has to be declared in writing, which in this case does not include electronic form (Art 98 (1) 4) of the TKG). The consent may be withdrawn at any time.

Art 98 of the TKG does not allow the collection and processing of location data for statistical purposes. This stipulation solely allows the transfer of this type of data for the purpose of offering a value added service to the extent that such data is necessary to provide the respective service.

As long as location data has to be considered personal data, there is no legislative authorisation in the TKG for collecting and processing such data for statistical purposes.

3.2.4.3. Telemedia Act (hereinafter the TMG)

The TMG incorporates the rules which constitute the transposition of the EPD into German law concerning personal data which is processed in connection with electronic information and communication services which are not telecommunication services exclusively consisting of the transmission of signals via telecommunication networks and which are not telecommunication-based services. In the relevant case of tourism statistics the TMG does not apply as long as the location data is being processed by a telecommunication service provider. The TMG, however, would be applicable if the location data originates from another source, such as provision of internet services, even if the internet is been accessed via a mobile device.

3.2.4.4. Federal Statistics Act (hereinafter the BStatG)

The BStatG regulates the terms and conditions of the collection, processing, presenting and analysing of data concerning mass phenomena in order to provide statistics for federal purposes.

According to Art 5 (1) of the BStatG federal statistics shall principally be ordered by law or in certain cases by federal ordinance ('Rechtsverordnung'). Art 18 of the BStatG stipulates that the BStatG shall also be applicable in case of a statistical survey by the European Union, provided that the respective European law does not state otherwise. Art 19

of the BStatG enables the Federal Statistical Office to participate in statistical programs or the elaboration of statistics of the European Union or international organisations.

Regarding the processing of personal data in connection with a statistical evaluation Art 16 of the BStatG contains detailed confidentiality obligations. Such data has to be kept strictly confidential unless:

(a) the respective person has agreed to the contrary in writing;
(b) the data has been obtained from a source available to the public or from an official source named in Art 15 (1) of the BStatG if the obligation to provide information derives from a law ordering federal statistics;
(c) the data is combined by the federal statistics agency or a state statistics agency with data referring to other individuals and presented in a statistical conclusion;
(d) the data cannot be associated with the respective person.

Therefore the confidentiality obligation does not apply if the information cannot be attributed to an individual person, hence, if the data is anonymised. According to Art 21 of the BStatG the combination of data from statistics with other information in order to re-identify individual persons for other reasons than statistical purposes or any other purpose accepted by the federal law ordering the statistic is forbidden.

The transfer of personal data between persons or entities appointed with the preparation of federal statistics is allowed as far as necessary for such purpose. The owner of the data can be obliged to cooperate and provide the requested data by law (Art 15 of the BStatG). According to Art 17 of the BStatG the relevant persons have to be informed about the purpose of the statistics and about their statutory rights in this regard.

3.3. Overview of Main Issues of Concern

3.3.1. Data Necessary for the Purposes of Tourism Statistics

Following initial data is required or desirable for the purposes of conducting tourism statistics on minimum level:

- the International Mobile Subscriber Identity (IMSI);
- the date and time of the start of the communication;
- the location label (*cell_id*) at the start of the communication;

- data identifying the geographic location of cells by reference to their location labels (*cell_id*) during the period for which communications data is retained.

All the above listed data constitute the data subject to the data retention obligation under the DRD (see Section 3.1.3.1.8). Hence, the required data must be retained by the MNOs as set out by each local law. Despite the obligation to retain the data (note the exception of Germany described in Section 3.2.4.2) such data, if deemed personal data, may only be processed in any other way if the data subject has given his prior consent or if there is a specific statutory basis for such processing or the data is anonymous.

3.3.2. Anonymous and Aggregated Mobile Positioning Data as Personal Data

As explained in Section 3.1.3.1.1 any data is considered personal data and is therefore within the scope of the DPD in case such data relates to an identified or identifiable natural person. The criteria of identifiability are further addressed in Section 3 of the Working Party Opinion 4/2007 On the Concept of Personal Data of 20 June 2007.

Moreover, the Working Party has stated in their Opinion on the Use of Location Data with a View to Providing Value-added Services as of November 2005 that since location data always relates to an identified or identifiable natural person, it is subject to the provisions on the protection of personal data laid down in the DPD. Moreover, Section 4.1.1 of the Geolocation Opinion stipulates that since location data derived from base stations relate to an identified or identifiable natural person, it is subject to the provisions on the protection of personal data laid down in the DPD.

Therefore, based on the Working Party opinions it should be concluded that location data is always considered to be personal data, and therefore within the scope of the DPD.

Nevertheless, it must be taken into account that the Working Party opinions are of advisory rather than compulsory nature. Moreover, it can be argued that the position of automatically treating all location data as personal data is not fully correct.

3.3.2.1. Aggregated Data

Based on information provided as the basis for carrying out this analysis it is presumed that the aggregated data cannot be traced down to an identifiable person at any time in any way. If such presumption is correct, it can be concluded that the processing (incl. collecting, storing, using, transmitting, etc.) of the aggregated mobile positioning data does not fall

within the scope of the DPD and is therefore freely usable without any data protection implications by persons who obtain possession of such data in aggregated form. This applies to the data brokers and other third persons that obtain the data in aggregated form from the MNOs.

It must be taken note of, however, that at the time of collecting the underlying data by an MNO the data is not yet in aggregated form. Therefore, collection and processing (rendering into aggregated form) of the data that can be traced down to an individual (a subscriber or user) by an MNO is deemed processing of personal data. Such processing by an MNO must therefore be conducted in accordance with the applicable laws.

It could be argued that if an MNO processes for the mere purpose of rendering into aggregated form of personal data that such MNO has in its disposal anyway such processing by the MNO should not be deemed to interfere with the relevant data subjects' privacy more than it would without said processing for such specific purpose. Certain representatives of the Estonian Data Protection Authorities have in course of verbal consultations opined the same. It must be taken into account, though, that formally there is no basis arising from the law that would support such interpretation. Therefore, in order for an MNO to act diligently, it should obtain due prior consents from relevant data subjects for the processing of their personal data for the described purpose. Given that obtaining the consents separately immediately prior to commencement of relevant data processing is not practically feasible, it is recommended that respective consent is given by the client in the client agreement concluded with the MNO.

3.3.2.2. Anonymous Data

The concept of anonymous data is somewhat more complex. It is understood that anonymous data may in principle be tracked down to an identifiable person although such possibilities are extremely limited. It needs to be clarified if such tracking down can be considered direct or indirect identifiability in light of the DPD and relevant local laws. The Working Party states in its Opinion 4/2007 on the Concept of Personal Data that "/…/ in general terms, a natural person can be considered as 'identified' when, within a group of persons, he or she is 'distinguished' from all other members of the group. Accordingly, the natural person is 'identifiable' when, although the person has not been identified yet, it is possible to do it (that is the meaning of the suffix '-able'). This second alternative is therefore in practice the threshold condition determining whether information is within the scope of the third element".

In its Geolocation Opinion the Working Party has stated that '/…/ smart mobile devices are inextricably linked to natural persons. There is usually direct and indirect identifyability. First of all, the telecom operator providing GSM and mobile internet access usually has a register with the name, address and banking details of every customer, in combination with several unique numbers of the device, such as IMEI and IMSI. Secondly, the purchase of extra software for the device (applications or apps) usually requires a credit card number and thereby enriches the combination of the unique number(s) and the location data with directly identifying data. /.../ Indirect identifyability can be achieved through the combination of the unique number(s) of the device, in combination with one or more calculated locations. Every smart mobile device has at least one unique identifier, the MAC address.'

It results from the above that any location data is personal data even if the relevant data subjects (MNO's customers) have not been but could in principle be identified based on such location data. Therefore, one should conclude that anonymous data (as opposed to aggregated data in the context of the study), to the extent a person can be singled out based thereon, may only be processed by an MNO upon the customers' prior consent or on a basis set out by the law. Such conclusion is not affected by the fact that both the DPD and EPD use the term 'anonymous' when referring to the state the data must be brought into for the purposes of the data controller being able to process thereof without the data subject's consent. In the context of the DPD and the EPD 'anonymous data' is meant as data based on which a person cannot be singled out.

To conclude, the applicability of the DPD and the relevant local laws depends on whether the mobile positioning data constitutes personal data or not. In case the mobile positioning data does not relate to an identified or an identifiable individual and is of purely statistical nature, such data is not personal data and the DPD and the PDPA (and local laws implementing the DPD in their respective legislations) do not apply.

3.3.3. Processing of Personal Data for Official Statistics Needs as a Statutory Basis for Processing Personal Data

If a data subject's consent is not obtained personal data may be processed only when a statutory basis for processing exists. Such statutory basis must be expressly set out by the law.

There are two possible bases whereon personal data may potentially be processed for the purposes of official statistics.

First, the general statutory basis for the processing of personal data arises from Art 7 (e) of the DPD (processing data for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data is disclosed). The prerequisite to applying this basis is that the state statistics authorities need the data in order to perform an official task given to them by law. Provided that the statistics authorities can request personal data from an MNO on such statutory basis, the MNO is obliged to transfer such requested data.

Second, the DPD also stipulates a more specific statutory basis in its Art 6 (b) under which Member States shall provide that personal data must be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards.

The implementation of Art 7 (e) and Art 6 (b) of the DPD in the investigated jurisdictions is varied.

3.3.3.1. Estonia

The OSA provides that a producer of official statistics may use personal data on the basis of and pursuant to the procedure provided for in the PDPA whereas it is not required to inform data subjects of the use of their personal data in producing official statistics. Therefore the OSA does not set out an individual basis for the state statistics authorities to process personal data for statistics purposes and the latter has to be done according to the PDPA as the general law.

Art 10 (2) of the PDPA transposes Art 7 (e) of the DPD into Estonian law. Thereby an administrative authority shall process personal data only in the course of performance of public duties in order to perform obligations prescribed by law.

Furthermore, Art 14 (1) 1) and 2) of the PDPA stipulate that processing of personal data is permitted without the consent of a data subject if the personal data are to be processed on the basis of law or for performance of a task prescribed by an international agreement or directly applicable legislation of the Council of the European Union or the European Commission.

The Estonian Data Protection Authorities have in course of verbal consultations expressed an opinion that Statistics Estonia should in principle be allowed to process personal data on this ground rather than on the grounds of Art 16 of the PDPA. It should be noted,

though, that Art 10 (2) and Art 14 (1) 1) and 2) are of very general nature and do not specify the terms and conditions of such processing.

Art 16 of the PDPA transposes Art 6 (b) of the DPD into Estonian law. As opposed to Art 10 (2) and Art 14 (1) 1) and 2) Art 16 sets forth terms and conditions of such processing in further detail. Art 16 (1) stipulates that data concerning a data subject may be processed without the consent of the data subject for the needs of official statistics only in coded form. Before handing over data for processing it for the needs of official statistics, the data allowing a person to be identified shall be substituted by a code. Decoding and the possibility to decode is permitted only for the needs of additional official statistics. The controller of the personal data shall appoint a specific person who has access to the information allowing decoding.

The PDPA does not specify what a coded form of data means. The underlying EU directives do not use the term 'coded form'. It remains ambiguous where one should draw a line between the 'coded' data and the data 'in a format which does not enable identification of the data subject'. We recommend holding further consultations with the Estonian Data Protection Authorities in this respect for clarification purposes.

Therefore, if an MNO as the data controller has collected its customers' location data it may transfer such to a government institution performing the function of producing official statistics even without the relevant data subjects' consent given that the data is in coded form. It does not follow clearly from the referred provision of the PDPA if the person that the MNO must appoint as the one having access to the decoding information must be an employee of the data controller or can also be a third person. Taking into account the general principles of personal data processing the relevant provision should be construed conservatively to be understood that only an employee or other person acting under the control and supervision of a data controller can be granted access rights to the decoding data.

Art 16 of the PDPA further stipulates that processing of data concerning a data subject without the person's consent for official statistics purposes in a format which enables identification of the data subject is permitted only if, after removal of the data enabling identification, the goals of data processing would not be achievable or achievement thereof would be unreasonably difficult. In such case, the personal data of a data subject may be processed without the person's consent only if the person carrying out the scientific research finds that there is a predominant public interest for such processing and the volume of the obligations of the data subject is not changed on the basis of the processed personal data and the rights of the data subject are not excessively damaged in any other manner.

Although pro and contra arguments can evidently be found, it would be difficult to argue that processing of location data for the purposes of statistics producing reasons is a predominant public interest if the same result can be reached by alternative means (i.e. alternatively collected data).

It should be further borne in mind that the processing of personal data for official statistics purposes without the consent of the data subject is permitted if the data controller has taken sufficient organisational, physical and information technology security measures for the protection of the personal data, and if such processing involves processing of sensitive data, has registered the processing of sensitive personal data. The processing of such personal data can only be commenced if the Estonian Data Protection Authorities have verified compliance with the requirements set out in the law and, if an ethics committee has been founded based on law in the corresponding area, has also heard the opinion of such committee.

Collected personal data may be processed for the purposes of official statistics regardless of the purpose for which the personal data was initially collected. Personal data collected for official statistics may be stored in coded form for the purposes of using it later for scientific research or official statistics.

Based on the Estonian statutory regulation described above it must be concluded that it is not clear whether Art 16 of the PDPA serves as a specific rule in relation to the more general Art 10 (2) and Art 14 (1) 1) and 2). If it does, the processing of personal data by statistics authorities must be conducted pursuant to the provisions of Art 16. In case it does not, it remains unclear what terms and conditions the statistics authorities must adhere to upon processing data under Art 10 (2) and Art 14 (1) 1) and 2) as well as in which case does Art 16 apply. In course of verbal consultations the Estonian Data Protection Authorities have indicated that in their opinion Art 10 (2) and Art (14 (1) 1) and 2) should be a sufficient basis for such processing. No further guidance was given though.

In case Art 10 (2) as well as Art 14 (1) 1) and 2) are to be considered an independent basis of processing data for statistics purposes, Statistics Estonia must be subject to a statutory obligation to collect and otherwise process the mobile positioning data to the extent relevant. Such obligation must arise from the law, an international agreement or directly applicable legislation of the Council of the European Union or the European Commission. The law does not clarify to what level of detail the obligation of Statistics Estonia should be set out by the law to be eligible under Art 10 (2) and Art 14 (1) 1) and 2) of the PDPA. Nor does the law specify based on which criteria it should be assessed whether specific data is necessary for the

performance of a specific task (e.g. whether the data should be inevitably necessary for the performance of the task or is it sufficient if the performing of a task is easier, more efficient, etc. as a result of processing the data whereas the task could in principle be performed by taking alternative measures as well). The Estonian DPA has opined in course of verbal consultations that they do not expect the law should necessarily set forth an expressed obligation to collect mobile positioning data from MNOs or similar detailed obligation. Therefore in principle, the general obligation of Statistics Estonia to produce official statistics arising from the OSA should be sufficient if Statistics Estonia provides sufficient arguments that it is not possible to produce required statistics without relevant location data.

Additionally to the obligation to perform official statistics arising from the OSA, Art 1 of the Regulation (EC) 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 (hereinafter the Regulation 692/2011) stipulates that the Member States shall collect, compile, process and transmit harmonised statistics on tourism supply and demand, and elaborates on such obligation in its further provisions. Under Art 8 (b) of the same it can be concluded that the MNOs' databases of positioning data are acceptable sources of data for official statistics. The regulations of the European Parliament and the Council are directly applicable legal acts in the Member States. Therefore, the provisions of the Regulation 692/2011 need not be transposed into Estonian law to be applicable locally. Therefore, the conclusion drawn herein supports the position that if Statistics Estonia needs the mobile positioning data in order to fulfil its task of producing official tourism statistics it may claim such data from the MNOs for processing thereof for such purpose. The latter conclusion is valid on the precondition that the data categories and intended use of the data is covered by the purposes determined in Art 3.1 of Regulation 692/2011. If those purposes do not require the use of positioning data the Commission could, according to Art 3.2 and 11 of Regulation 692/2011, adopt a delegated act in order to adapt this list of subjects and characteristics of the required data. With such act the commission could further extend the purposes included in Art. 3.1 of Regulation 692/2011 on statistical purposes, which require as source the MNOs' databases of positioning data.

It is more ambiguous, though, on which terms and conditions Statistics Estonia has to process the data, among other, whether the data should be aggregated or made anonymous at a certain point of time.

The Regulation (EC) 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical

confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (hereinafter the Regulation 223/2009) sets out the rules of confidentiality of statistical data. 'Confidential data' in the context of the said Regulation means data which allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit. Therefore, the identification criteria of a statistical unit are, by large, similar to the criteria of identification of a data subject pursuant to the DPD. Under Art 3 6. a statistical unit may, among other, be a natural person or a household. Thus, it can be concluded that in principle the confidential data under Regulation 223/2009 may, in case of a natural person being the statistical unit, be data by which the natural person can be identified, i.e. personal data in the meaning of the DPD.

Pursuant to Art 20 2. of Regulation 223/2009 confidential data obtained exclusively for the production of European statistics shall be used by the national statistics authorities and other national authorities exclusively for statistical purposes unless the statistical unit has unambiguously given its consent to the use for any other purposes. Therefore, presuming that the confidential data in light of the Regulation can be deemed personal data, one possible conclusion would be that the authorities may process the personal data for statistical purposes without the data subject's consent. It needs to be taken into account though that Art 20 2. of Regulation 223/2009 applies to confidential data obtained exclusively for the production of European statistics. Therefore, it is questionable whether this applies to data that was originally collected by a third person (an MNO) for its business purposes and only thereafter transferred to be processed for statistical purposes.

The national statistics authorities must take all necessary measures to ensure the harmonisation of principles and guidelines as regards the physical and logical protection of confidential data. Those measures shall be adopted by the Commission in accordance with the regulatory procedure referred to in Article 27(2). Furthermore, officials and other staff of the relevant authorities having access to confidential data shall be subject to compliance with such confidentiality, even after cessation of their functions. Any transmission of confidential data by the national statistics authorities may only be conducted according to Article 21 of the Regulation 223/2009.

As far as the transfer by MNOs of personal data to Statistics Estonia is concerned, Art 14 (2) 1) of the PDPA stipulates that communication of personal data or granting access to personal data to third persons for the purposes of processing is permitted without the consent of the data subject if the third person to whom such data are communicated processes the personal data for the purposes of performing a task prescribed by law, an international agreement or directly applicable legislation of the Council of the European Union or the European Commission. Thus, provided that Statistics Estonia needs such data in order to perform its official tasks under the OSA and the Regulation 692/2011 the MNOs are entitled and obliged to transfer such personal data to respective authorities.

In case the business model used involves transferring personal data by an MNO to a data broker or mediator rather than to Statistics Estonia directly, a question arises if such transfer of the data, even if eventually used for official statistics purposes, qualifies as lawful data transfer by an MNO. In order to eliminate the risk of the data brokers/mediators of illegal processing of personal data they should also be subject to performing a statutory task commissioned to them under the law or a contract under public law concluded with the state statistics authorities.

3.3.3.2. Finland

There are two main issues of concern related to obtaining and using mobile positioning data for the purpose of producing tourism statistics in Finland.

First, the current Statistics Act obliges enterprises to provide data on their products and services but not on the clients consuming these products and services. Therefore the current Statistics Act does not oblige MNOs to provide mobile positioning data to Statistics Finland.

The second key question is whether anonymous raw data provided by MNOs constitutes personal data or not under the HTL. The Data Protection Ombudsman, the main data protection authority in Finland, has been consulted regarding this question and they are currently preparing a statement where this question will be addressed. If the result of this consultation is that the raw data, although pseudonymous, still constitutes personal data, there arises a need for a mediator to process the data into aggregated form on the premises of the MNO. Such mediator has to have access to the raw data maintained by the MNO as well as possess methodological competence and tools to process the raw data into aggregated data. This scenario involving a mediator is challenging technically, financially and in terms of organisation.

3.3.3.3. France

The prerequisites for an MNO to be able to transfer location data deemed to be personal data to the statistics authorities would be the following.

Generally, for the transfer of personal data to the statistics authorities by an MNO, the MNO must itself lawfully process the data and respect all of its obligations as data controller. The data cannot be transferred by an MNO for statistics purposes without the prior consent of the data subject. Moreover, the MNO must have duly filed with the CNIL the notification corresponding to its data processing, mentioning the transfer of data to third parties, and must provide to subscribers all mandatory information regarding such data processing.

The transfer agreement entered into between the statistics authorities and the MNO should provide that the MNO complies with the above. A copy of the MNO's notification(s) should also be requested.

Pursuant to Art 34-1.VI of the PECC, the MNO can only transfer the identification of the subscribers, the technical characteristics of the subscribers' communications and the location data.

As the statistics authorities should be considered as being data controllers, the statistics authorities also need to file for a notification as mentioned above and respect all of the obligations provided under the ITLA for data controllers.

As the data will be used for statistical purposes, it would be useful to verify whether the data will be anonymised shortly after being transferred to the statistical authorities. In such a case, the CNIL should be contacted to verify that the anonymisation procedure does comply with the ITLA. The MNO would then only need to mention in its agreements with subscribers that their data will be transferred to the statistical authorities. If this is not the case, all of the information indicated in Section 3.2.3.1 should be provided by the MNO in such agreements.

As an example, the CNIL considered in 2010, that devices analysing the attendance of certain places (such as airports or commercial centres) by capturing data transmitted by mobile phones and therefore calculating the geographical position of data subjects were compliant with the ITLA if a clear information was displayed in such places on the purpose of the devices and the data controller, and the data was anonymised shortly after being collected by an anonymisation process controlled and approved by the CNIL.

In addition, as for Estonia, if the location data is collected to comply with Regulation 692/2011 and the location data is considered as falling under Art 8 (b) of the Regulation

692/2011, the statistical authorities may be considered as acting within the framework of a legal obligation or the performance of a public service mission. In such a case, under Art 6 of the ITLA, consent of the data subject is not necessary for the processing of the location data. As for Estonia, the data will also need to be treated in compliance with Regulation 223/2009.

However, we strongly recommend that the opinion of the CNIL be formally requested on this issue, in particular since location data may not be necessary to provide the information requested under Regulation 692/2011 (in particular pursuant to Art 3 of the Regulation 692/2011) and Art 34-1.V of the PECC, which requests the consent of the data subject for the use of location data, does not provide for any exception in relation to compliance with legal obligations or performance of a public service mission.

Transfer of location data that is personal data to a third party data broker must be conducted on the following terms.

For the purposes of the following analyses it is assumed that the third party data broker would be a data processor and would be situated within the European Union.

The CNIL considers that the following criteria should be taken into account to verify whether a services provider acts as data controller or data processor:

(a) whether the services provider receives general or detailed instructions from the customer,
(b) the level of control exercised by the customer over the services and the data transferred to the services provider (e.g. whether the service provider has a right to use the data as it sees fit),
(c) whether the services provider acts under the customer's name or its own name and/or re-uses the data for its own purposes,
(d) the level of expertise of the services provider (e.g. whether the technical means used to provide the services are imposed by the service provider and cannot be modified by the customer either because the customer does not have the necessary skills or because the software does not require any specific developments).

If the third party data broker is a data processor, it is assumed that it would act on behalf of the statistical authorities. A data processor is not subject to the obligations provided under the ITLA. However, Art 35 of the ITLA provides that the processor shall offer adequate guarantees to ensure the implementation of the security and confidentiality obligations

incumbent upon the controller. This requirement does not exempt the data controller from its obligation to supervise the observance of such measures.

The contract between the processor and the data controller must specify the obligations incumbent upon the processor as regards the protection of the security and confidentiality of the data and provide that the processor may only act upon instructions from the data controller.

If the third party broker is a data processor, consent from the MNO's customers is not necessary for the data to be transferred to it, as the data processor acts under the control of the data controller. A contract between the statistical authorities and the third party broker will, however, need to be entered into, providing for the obligations mentioned in the preceding paragraph.

3.3.3.4. Germany

German law allows the processing of data for statistical purposes if a German or European law enables the Federal Statistical Office or a comparable European Authority to do so. The main requirement in this case is that the personal data is kept confidential, unless the data cannot be associated with the individual mobile phone user, i.e. by anonymising the data and combining data referring to several users in a statistical conclusion. However, a law enabling the Federal Statistical Office to claim and process location data does not currently exist in Germany.

Provided that such law was enacted, the Federal Statistical Office could claim the transfer of such location data from any person incl. the MNOs. According to Art 17 of the BStatG the relevant persons have to be informed about the purpose of the statistics and about their statutory rights in this regard. The further conditions for the claim of location data would be determined in the respective legal provisions and, therefore, cannot be foreseen at this point of time.

According to the TKG location data which can be qualified as traffic data can only be stored by an MNO as long as storing such data is necessary for the purposes of invoicing for the services, provision of an itemised bill, the identification and elimination of faults and malfunctions, identification of misuse and call tracing in case of telephone stalking (Art 96 to 101 of the TKG). During this period the MNO could be obliged by respective law to transfer these personal location data for statistical reasons to the Federal Office for statistics.

If, however, an MNO stores location data for a longer period than needed for the above mentioned purposes, it has to anonymise the location data since the processing of such personal data is only allowed for the purposes described in Art 96 and 98 of the TKG.

The Regulations 692/2011 and 223/2009 are directly applicable in each member state of the European Union. Therefore, the conclusions drawn in Section 3.3.3 concerning the effect of these two regulations on the Estonian legal system are valid for the German jurisdiction as well.

Nevertheless it is highly recommended that the German Data Protection Authorities are consulted regarding whether the anonymised location data still constitutes personal data or not, if the establishing of a connection between this data and a certain person is theoretically possible but, due to the complexity and effort needed, effectively very improbable.

3.4. Conclusion and Recommendations

The above overview lists and introduces the relevant acts on the level of the EU as well as the four sample Member State jurisdictions covered by the study. Although the EU directives have been transposed and implemented into local laws (with the exception of the DRD not being implemented in Germany as explained above), the implementation practice has proved to be somewhat different resulting in variety of approaches concerning the topic in question. Additionally, the practice applied by regulators in each country, if existing at all, differs to quite some extent. Based on the above, the following issues are highlighted.

3.4.1. Relevant Data

The data relevant in terms of the study (Section 3.3.1) corresponds to the data subject to the data retention obligation under the DRD. It is essential to note that the MNOs' obligation to retain data under the DRD does not exempt the MNO from processing personal data according to the general rules of the personal data processing arising from the DPD and the EPD. To further extent the DRD is not relevant in the context of the study since the obligation to retain data arising thereunder is imposed for the purposes of making data available to law enforcement authorities for the purposes of the investigation, detection and prosecution of serious crime.

3.4.2. Location Data as Personal Data

Conclusion: It is essential to identify whether the data provided by an MNO is deemed personal data. The applicability of the DPD and the relevant local laws depends on whether

the mobile positioning data constitutes personal data or not. The DPD does not apply in case the mobile positioning data processed does not qualify as personal data. In such case the data could be freely used, including transferred to the state statistics authorities or third party service providers as may be necessary.

In general, personal data means any information relating to a natural person who is or can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. Such principle derives from the DPD and is implemented in Member States' local laws with slight differences as to which means one should take into consideration when determining if a person is identifiable or not. In France, for example all the means that the data controller or any other person uses or may have access to should be taken into account. In Germany, means that require a disproportionate amount of time, expense and effort need not be taken into account. However, all in all a conclusion can be drawn that conservative and strict approach is to be taken upon construing the concept of personal data which is why the following can be concluded.

Aggregated data is not subject to the provisions of the DPD and can therefore be processed without limitations arising therefrom. The issue is outstanding in terms of data that is not in aggregated form. Since the qualification of data as personal data comes down to a matter of assessment in each given situation, there is no universal answer to that question. The laws and the practice in each sample jurisdiction are slightly different. Out of the four sample countries France is the only one where the regulator has provided some guidance as to anonymisation keys (refer to Section 3.2.3.1 for details). In Finland the Data Protection Ombudsman has stated that purely pseudonymised data is not sufficient to be considered as anonymous. Today's guidance on the Working Party level has not addressed the topic at hand in sufficient detail. The general approach of the Working Party is, though, that mobile positioning data is as a rule personal data. Therefore, if not in aggregated form, it should be presumed that location data is personal data.

Recommendation: Updating and specifying the Working Party guidance would be one option to clarify the issue and achieve unified approach throughout the EU. However, given the differences in transposing the EU directives into local laws and consequent differences in implementation may prevent giving a unified guidance.

As long as the local laws differ, giving further guidance by local regulators (similarly to what the CNIL has given in France) on implementation of the statutory rules is recommended. If guidance on anonymisation keys is given, it may, depending on such

guidance, occur that in certain cases anonymised location data is not deemed personal data and the processing thereof is subject to less formalities (e.g. if the anonymisation keys are valid for a sufficiently short periods of time, etc.). It is also highly recommended to consult the data protection authorities of all relevant EU countries to clarify the accepted practice of applying local laws.

3.4.3. Processing Location Data for Official Statistics Purposes

Conclusion: If location data is in aggregated form (i.e. cannot be used to single out any individual) it can be processed for statistics purposes without limitation from the data protection perspective. If location data is deemed personal data (i.e. if it is not in aggregated form) the approach differs from country to country as to what specific terms and conditions must be met in order for such processing to be lawful.

In Estonia it is most likely possible to draw a conclusion that personal data may be used for official statistics purposes without the data subject's consent if certain preconditions (see Section 3.3.3.1) are met. However, it is not fully clear if the general basis for processing for the purposes of performing an official task by the statistics authorities would be sufficient to render such processing without the data subject's consent lawful or more specific terms have to be adhered to.

In Finland the effective law does not oblige MNOs to provide mobile positioning data to statistics authorities. Unless the legislation is changed to specifically require MNOs to provide the data, MNOs would have to process the data on their premises and render the data into aggregated form before providing it to statistics authority.

In France the data cannot be transferred by an MNO for statistics purposes without the prior consent of the data subject. The law requires consent for the processing of location data. Further, it is unclear as to whether compliance with Regulation 692/2011 is considered as a condition which entitles the statistic authorities to collect data without the data subject's consent. Even in such a case, this does not necessarily mean that an MNO would have an obligation to provide this data to the statistic authorities.

The German law does not currently stipulate the right of the statistics authorities to request relevant data from the MNOs. Such law, if enacted, would have to set forth the detailed terms and conditions of processing the personal data for statistics purposes. However, the Regulation 692/2011, on the preconditions described in Section 3.3.3, would entitle the official statistics producer to claim data from MNOs and process thereof.

Therefore, although there are regulations in place on general level in all jurisdictions, little clarity as to the terms and conditions of processing location data lies in details. In each jurisdiction the principles of the EU directives have been implemented differently. Furthermore, the legal practice and interpretation of the local regulators differ to quite some extent. Therefore the obstacles arising from each relevant jurisdiction are also different.

Recommendation: Although the EU directives serve as the basis for relevant regulation in the Member States, the directives provide for quite some flexibility to the Member States upon transposing the directives into local laws. Consequently, the general unique principles set forth by the EU have been implemented differently in detail in local laws. Thus, in order to obtain a full overview of the obstacles set by local laws, the local laws of all relevant Member States need to be analysed in detail. Due to the variety of issues across the investigated sample jurisdictions (and presuming the jurisdictions not investigated for the purposes of this study each have their own difficulties of implementation) it is not possible to draw unique and comprehensive conclusions as to how local laws should be amended in order to eliminate all obstacles in processing the location data for statistics purposes. Local laws of each jurisdiction should be reviewed and revised where necessary if no common obligatory guidance is given on the EU level.

One possible solution to eliminate the need to obtain data subject's consent for the processing of personal data for statistics purposes is to verify if the data categories and intended use of the location data in the context of the study would be covered by those set out in Art 3.1 of Regulation 692/2011 that is directly applicable in all Member States. If not, the list of purposes of use could be modified by a delegated act of the Commission as further described in Section 3.3.3. It is not clear, however, if this would be a viable solution for e.g. in France and Finland although it most likely would be in Estonia and Germany.

Notwithstanding the above, since the practice of interpreting and applying the relevant acts is limited and ambiguous in the investigated countries it is highly recommended to further consult the regulatory authorities in each relevant Member State.

Alternatively, the legislation on the EU level should be amended to be introduced in sufficient detail with more detailed guidance to the Member States for implementation thereof. Given that the Draft Regulation is currently in the process of being negotiated, it would be useful to identify the possibilities of introducing such clarifying provisions in the Draft Regulation that will be directly applicable in the Member States. More specifically, since the delegated acts that the European Commission will be entitled to give under the Draft Regulation will likely be the instruments whereby the specific criteria of processing data for

statistics purposes will be set forth, the focus should be on finding ways to highlight the current deficiencies to the Commission and help them draft respective delegated acts so as to introduce the specific enough provisions. That would preclude the need to amend each relevant local law.

Apart from the legislative measures, contractual measures can be applied to overcome the statutory obstacles. The main issue common to all investigated jurisdictions is that either it is not possible to transfer the personal data to statistics authorities without the data subject's prior consent or it is not fully clear if it may be done and on what conditions. One should bear in mind that obtaining separate consents from data subjects is not practically feasible. Therefore, in each case where the data subject's consent for transfer of data to by an MNO to the state statistics authorities is needed, it is recommended that the MNOs' agreements with their customers include the customers' respective consent. The consent should not be part of the standard terms and the refusal by a customer to give one should not deprive the customer from entering into the agreement with an MNO. The data subject should be able to revoke the consent at any time. Detailed requirements to the terms and form of the consent should be investigated in each relevant jurisdiction.

3.4.4. Transfer of Location Data to Third Party Data Brokers

Conclusion: The transfer of location data that is personal data to third parties acting as data brokers or in similar function can only be done upon data subject's consent. The only exception to this is the transfer of data to a third party performing an official task as the representative of the state statistics authorities. In order for the processing by a third party to be lawful, the relationship between the statistics authorities and such third party should be designed pursuant to the requirements of each local jurisdiction. In Estonia, for example, data brokers should be commissioned with an official task by the statistics authorities under the law or a contract under public law.

Recommendation: In the event third party data brokers are used in the model of processing the location data for statistics purposes, and provided that the conditions to the lawful processing by the statistics authorities are met in each relevant jurisdiction, it is necessary to identify what type of legal relationship needs to be created between the statistics authorities and the data broker in each relevant Member State jurisdiction.

4. Technological Opportunities and Barriers

The current chapter discusses the technological aspects from theoretical and practical points of view of retrieving data from MNOs and processing it to generate usable tourism statistics. Legal resolutions and financial resources have a major impact on the actual technological solution depending on where the data can be processed and how. The current assessment is based on the practical experience of retrieving data from MNOs as well as the knowledge from surveys and interviews with experts and stakeholders.

The survey respondents raised questions about the solution to privacy issues – to work on an aggregate scale or to have robust methods allowing the anonymisation of the data. The questions by the respondents were: Is aggregated or fully anonymised data less sufficient for statistics use? Does the statistics have to be based on exhaustive data or is a sample sufficient? In the end, most of the worries for both users and MNOs were about the methodological, rather than the technological aspect of access and use of data. The technology has matured enough for the MNOs to be confident in their ability to provide necessary datasets. What needs to be considered is the cost of technology related to large databases, and also what, when and how often one measures. Then, it comes down to processing of data and finding the necessary algorithms, which can prove to be difficult.

This report analyses the part of the data extraction process starting from logical data movement flows within a single MNO's network (Public Land Mobile Network – PLMN) core systems up to the start of processing the data specifically for tourism statistics purposes. For reasons of simplification the discussion will only be on the meta information of the events, as actual data exchange (voice communication, content of messages and internet traffic) is not within the interest or scope of the current study and can be considered much more sensitive in terms of the privacy of the subscribers.

There are several issues that need to be addressed before the appropriate data can be retrieved from an MNO:

1. Following all privacy and data protection laws to protect subscribers' identities;
2. Identifying the initial source (inbound, domestic, outbound) and specifications of the initial raw data;
3. Handling the geographical references of the raw data;
4. Preparing the data (data cleansing) before specific tourism statistics processes begin and tunnelling the packages to a tourism statistics processor.

Some of these problems directly influence the methodology used to process the data further and are therefore discussed more thoroughly in Report 3a of the current study.

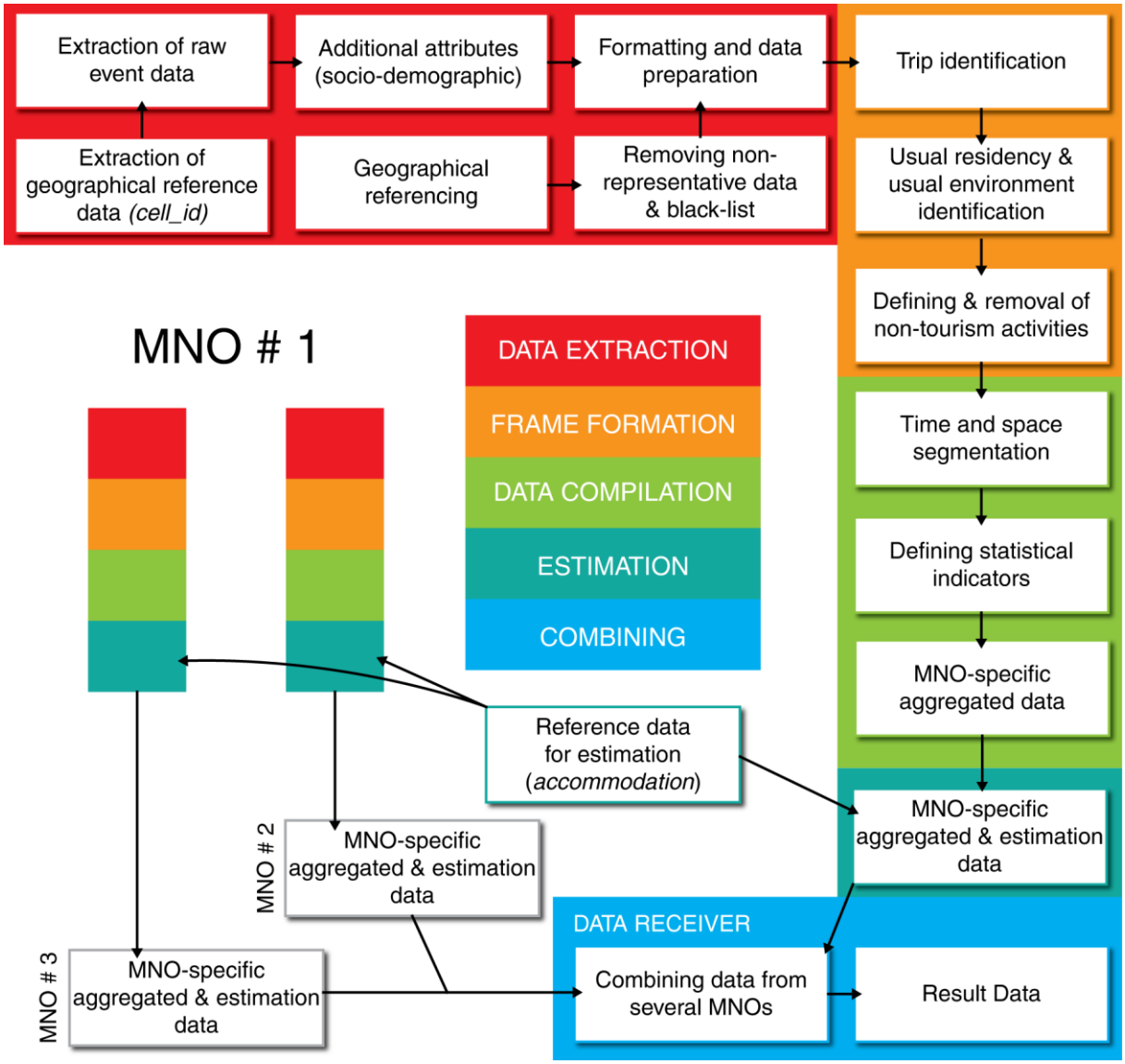

Figure 4. Data processing steps.

Figure 4 represents the general steps of the process from the raw data to the final aggregated and estimated results. The process includes data extraction (covered in the current report), frame formation, data compilation, estimation and aggregation and final combination (covered in Report 3a).

4.1. Source for Initial Raw Data

There are two main technological methods for locating mobile phones from the infrastructure of MNOs: active and passive mobile positioning. Active mobile positioning (also referred to as MPS – Mobile Positioning System) is a technological process where a

mobile phone is traced in real-time using measurements from the network (e.g. trilateration of antennae through active ping for the phone in the network – similar to making a phone call without it actually taking place). This method is used by emergency services in determining the location of caller's mobile phone location; in several services by MNOs (locating a family member or friends, nearest point of interest); and also in research. The use of such method is technologically limited by the number of phones that can be located at the same time and by the requirement to have consent from the owner of the phone.

Passive mobile positioning is extracting data representing information about historical locations of the phones from the log files of MNOs. This method allows for a longitudinal view of all subscribers.

Although active mobile positioning can be more accurate and provide more precise location history of the phone (given that the phone is located continuously over a specific period of time for a long period), it is more intrusive for the system resources of MNOs and might affect the work of the main systems. The number of mobile phones that can be located using MPS is therefore limited (e.g. 10,000 phones at once). Passive mobile positioning usually does not require extra network resources like MPS does and can produce data for many more subscribers (all if a sample is not used), however the accuracy suffers (not antennae trilateration) and for some data source (i.e. CDR – Call Detail Records), the data is more chaotic and unevenly distributed in time compared to active positioning. Active mobile positioning is also more concentrated on a person (requiring consent from the user) whereas passive mobile positioning concentrates more on the bulk data.

Current study concentrates on the data from passive mobile positioning that is more suitable for tourism statistics. From a technological point of view, MNO systems are commonly similar because of the widespread standards in the telecommunications industry (Figure 5). However, there is some variability in specific technological solutions and therefore the essence of the initial data that can be extracted from MNOs can also vary. Initial raw data should contain the facts (events) represented by at least time, location and subscriber's identification attributes.

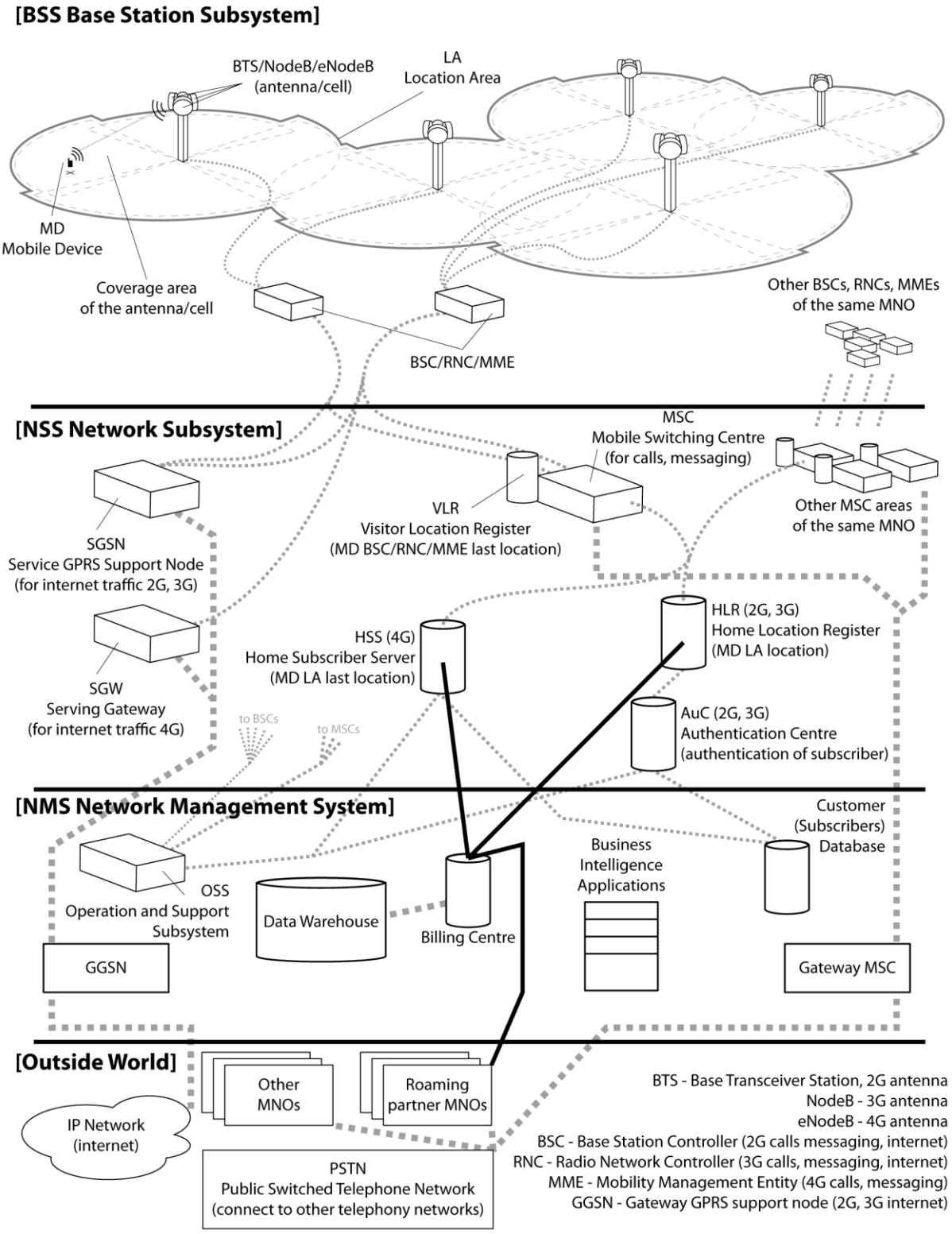

Figure 5. Simplified standard architecture of MNO.

Technically, the first question is the mapping and identification of the sources of data required from different registries and databases within an MNO's system. Three main types of

data can be distinguished based on their origin within the MNO's systems that are required for tourism statistics:

a) (any kind of) event (meta-)data about subscribers' activities from the MNO data stream;
b) geographical cellular (network) referencing data;
c) attribute data for the subscribers (e.g. demographic information from the customer database).

Most MNOs have many more internal custom systems, databases and registries that might be a part of the location data flow (e.g. security applications, network management, etc.) but as these are mostly custom, their use as the data source depends on the technical specifics of the MNO.

4.1.1. Event Data

Event data can be divided into internal and external network events and furthermore similarly to the forms of tourism statistics:

a) MNO internal events:
    a. inbound roaming;
    b. domestic;
b) MNO external events:
    a. outbound roaming;

As mentioned before, depending on the specific technical solutions within the MNO, there are several options for retrieving the raw event data of the subscribers. The main requirement is for the data to be stored periodically in a central database (can also be distributed in case of very large MNOs) where it can be extracted from. There can be several databases that store these events of which some might hold the event data in binary format (i.e. CDR data is originally stored in binary).

4.1.2. Inbound Roaming and Domestic Data Sources

Inbound roaming and domestic processes are almost identical within the MNO's network systems. There are several nodes in the network where event data is handled, processed and/or stored during the time when a mobile device is active within the MNO's network. On the BSS (Base Station Subsystem) level, the communication (and therefore the event metadata) between the mobile device and the network infrastructure (MD –

BTS/NodeB/eNodeB – BSC/RNC/MME, see Figure 5 for explanations) is usually very intensive during data transmission and almost all events are relayed to BSC/RNC/MME nodes. The MNO's central infrastructure does not always know the location of a mobile device within the network. This is why MSC sends location update requests to store the location information in the VLR (Visitor Location Registry) for better network management. Typical events that are handled within the BSS are.

a) switching the MD on;
b) switching the MD off;
c) receiving a call;
d) sending a call;
e) periodic update during the call;
f) termination of the call (by a MD or other party);
g) sending a message (SMS and MMS);
h) receiving a message (SMS and MMS);
i) initiation of the internet connection;
j) periodic update during internet connection;
k) termination of the internet connection;
l) handover (handoff) during the active phone call;
m) change of the active antenna (antennae in 3G and 4G) within LA;
n) any service communication between the network and MD (e.g. location-based services' active positioning of the MD by MPS);
o) LA (Location Area) change;
p) periodic LA update.

Very diverse information about the event is contained in the information metadata package, including (mostly but not always) information about the subscriber's identity and involved technical nodes (e.g. antennae ID/CGI – Cell Global Identity, physical communication parameters). The nature of the events and attributes handled depends on communication standards (2G, 3G, 4G) and is very extensive and complicated and therefore will not be covered in this report. Because of the massive data volume and traffic required, most of these events are deleted upon the new event update or are deleted periodically.

Although this depends on the specifications of individual MNOs, all events except change of active antenna within the LA are usually relayed further to the MSC (Mobile Switching Centre), the next logical node in the network. The MSC is responsible for communication between NSS and BSS levels for calls and messaging and is tightly connected

to the VLR (Visitor Location Registry), which stores the active location (BSC code) of the MD. This location is updated periodically and upon the occurrence of some events. Similarly to the MSC, the SGSN (Service GPRS Support Node in 2G and 3G) and SGW (Serving Gateway in 4G) are responsible for internet communication between the MD and the IP network (internet).

As with the BSC, most of the events are only stored within the MSC for a limited time and not relayed further to the central system of the MNO. Traditionally UDRs (Usage Detail Records) are sent to a Billing Centre or directly to the MNO's business intelligence applications from the communication between the HLR (Home Location Registry) and VLR via the MSC or SGSN/SGW. UDRs are typically essential for the billing purposes (retail and wholesale), network traffic management and monitoring, data warehousing, reconciliation system, fraud management, provisioning feed to sub-systems and other functions of the MNO. UDRs are mostly composed of CDRs (Call Detail Records – calls and messaging) and DDRs or IPDRs (Data Detail Records or Internet Protocol Data Records – both usage of the internet). Upon the occurrence of this data in the NSS, they are relayed to the Billing Centre and/or Data Warehouse and/or Business Intelligence Applications directly or via an event mediation system, where they are processed and stored according to technical configurations (some of which are derived from legislation – see 3.1.1.3 Directive 2006/24/EC). CDRs are considered the easiest initial source of event data as they are standard and exist in all MNOs. However, the event count per subscriber for CDRs depends largely on the behaviour of the subscriber.

As described earlier, there are several alternative options for extracting the event data from the network (see Table 1). Such alternatives will not be fully described in this report as they are all highly technical with large variations depending on configuration and licence agreements between the MNO and their technical platform providers. The MNOs of one country most likely do not have the same configuration and can provide different levels and qualities of the event data apart from CDRs.

Different sources can be combined totalling the average number of events per subscriber per day to 300-400. However such number most probably requires additional technological implementation from MNOs and will affect the requirements of the resources to be able to process such a large amount of data. Adequate balancing has to be made between implementation requirements (cost) on the MNO side on the one hand and for processing of the data, the number of the events that can be extracted per subscriber and the quality of the resulting statistics on the other.

Table 1. Description of possible data sources from an MNO for tourism statistics.

Data extraction option	Requirements	Description of the data
CDR data from Billing Centre/Data Warehouse/Business Intelligence Applications	Almost always available and easiest way to extract.	Events based on active usage of mobile phones – calls, messaging. Good for national level statistics for longer periods. Quality decreases with smaller time and spatial scales. MNOs usually do not require implementing any additional hardware to extract the data. However software and special hardware for storing and processing the extracted data for tourism is most likely required. * Number of events per subscriber per day: ~3...8.
Internet usage event (DDR or IPDR).	DDRs or IPDRs are not commonly stored by MNOs, but usually the addition of such data does not require very large technical investment on behalf of the MNO except for the licence of the platform provider.	DDRs or IPDRs increase the number of events by subscribers considerably (mostly domestic, less inbound roaming) but require much more storage (and sometimes network) capacity to be stored. The further processing of such dataset will increase the cost (time and resources) of processing and might not be cost-effective in terms of the increase of quality of the results compared to the cost of the processing. * Number of events per subscriber per day (not including CDRs): ~20...300+ for domestic, much less for inbound and outbound roaming because of high roaming prices.
Event data between BTS/NodeB/eNodeB and BSC/RNC/MME (Abis level)	Usually requires physical implementation of devices (probes) or designated hardware that sends Abis level information to the central system. Unless this is already implemented by the platform provider, this requires investment and a very large storage and traffic capacity to relay and store all the data.	Compared to CDRs, the number of events per subscriber will increase dramatically (both for domestic and inbound roaming). This requires much more storage (and sometimes network) capacity to be stored. The further processing of such dataset will increase the cost (time and resources) of processing and might not be cost-effective in terms of the increase of quality of the results compared to the cost of the processing. Because at the BSS level TMSI (see 4.1.4) is used instead of IMSI, a process of mapping TMSI to IMSI (or other subscriber identity) might be required upon the extraction of this data. * Number of events per subscriber per day (not including CDRs or DDRs): ~25...100.
Event data between BSC/RNC/MME – MCS	Unless this is already implemented by the platform provider, this requires investment in hardware, software and licence costs.	Compared to CDRs, the number of events per subscriber will increase substantially (both for domestic and inbound roaming). This requires much more storage (and sometimes network) capacity to be stored. The further processing of such dataset will increase the cost (time and resources) of processing and might not be cost-effective in terms of the increase of quality of the results compared to the cost of the processing. * Number of events per subscriber per day (not including CDRs or DDRs): ~15...35.
** The numbers of events are estimated and designed to provide the possibility to compare between different data sources. Actual numbers depend largely on the technological configuration of the network and many other factors including subscribers' behaviour).*		

The minimum number of attributes in any kind of event data usable for tourism statistics should include:

- the identity of the subscriber (including the identity of the country of origin, e.g. IMSI);
- start time of the event;
- reference to active antennae (CGI) at the start time of the event.

There are many more possible attributes that can be included in the event records. Some of them can be useful for further processing, but will not be discussed in the current report as there are simply too many options with little effect.

It is not possible to estimate the most suitable (cost-effective or burden-effective) data specification for generating tourism statistics. Obviously the more data there is, the better the results are, but the cost of implementation and data processing on the part of MNOs as well as the processing party (e.g. NSI) plays a significant role on the actual specification of the initial raw data to be used. Depending on the requirements for the quality and accuracy of the result data, different (and possibly hybrid) solutions should be used. The CDRs are the easiest in terms of burden and the required processing resources, but might be insufficient in small time and spatial scales. However if it is possible to add data concerning the initial (and last) appearance of the inbound roaming phone in the network, it would be a valuable addition describing the duration of the stay of the phone in the roaming service and would compensate potential under-coverage of the durations based on CDRs.

4.1.3. Outbound Roaming

Outbound roaming data for one MNO is inbound roaming data for its roaming partner MNO and is therefore handled internally (within the network of the roaming partner MNO) according to the specified technical configuration and cannot usually be influenced by the first MNO. The actual metadata exchange between MNOs is a subject of the legal and technical agreements and usually follows international standards. Therefore, even if the partner MNO produces much more event data on its side, only a portion of the events is provided for the first MNO. Those are usually events for billing, roaming management and monitoring purposes. In most cases, outbound roaming data comprises CDRs (incoming and outgoing calls, messaging) and DDRs or IPDRs (roaming internet usage). The data for outbound roaming should include:

- the identity of the subscriber;

- initial time of the event;
- a reference to the roaming partner (MNO) from which the country of roaming (MCC) can be established.

4.1.4. Subscriber's Identity Code

There are four initial identity codes that an MNO can use to identify the subscriber:

- IMSI – International Mobile Subscriber Identity;
- MSISDN – Mobile Subscriber Integrated Services Digital Network Number;
- IMEI – International Mobile Station Equipment Identity;
- Custom ID of the MNO.

Because the IMEI is the identity code for the mobile device, its use is discouraged as subscribers' identities can create bias (the same device used by different subscribers over a period of time). There is little information about the customer identity system within MNOs as they are often based on very different logics. Therefore, a customer's ID code will not be discussed more thoroughly in this report.

An IMSI/MSISDN/IMEI may occur:

- When a phone device is changed:
  - the IMEI changes;
  - the IMSI might change if the subscriber gets a new SIM;
  - the MSISDN should remain the same.
- When a subscriber changes his/her service provider within a country:
  - if the device is the same, then the IMEI remains the same;
  - the IMSI will change as a new SIM is provided;
  - the MSIDN will remain the same if the subscriber uses a number portability service available in most EU and many other countries.
- When a subscriber changes his/her service package within the same MNO:
  - the IMEI will remain the same if the device is the same;
  - the IMSI should remain the same unless a change of SIMs is required by the business logic;
  - the MSISDN remains the same.

Table 2. Change in IMSI and MSISDN (0—remains same, X—definite change, ? —possible change). See 4.1.5 for an explanation of the abbreviations.

Change	IMEI	IMSI			MSISDN		
		MCC	MNC	MSIN	CC	NDC/ NPA	SN
Same MNO, new device, same SIM	X	O	O	O	O	O	O
Same MNO, new device, new SIM	X	O	O	X(?)	O	O	O
Change MNO, same device, number portability	O	O	X	X	O	O	O
Change MNO, same device, no number portability	O	O	X	X	O	X	X
Same MNO, new service package, same device	O	O	O	?	O	O	O
Same MNO, same device, change of phone number	O	O	O	?	O	O	X

Although an MSISDN, at least in EU countries and countries with number portability regulations, is more permanent compared to an IMSI (see Table 4), an IMSI is the main standard identity code that is almost always included in the event records and data exchange between MNOs. By GSMA standards, an MSISDN might not be stored/included in many cases. Because of that using an MSISDN as the main identifier might require additional technical implementation for some MNOs. Therefore, an IMSI should be used as the basis for the unique identity code for the subscriber in inbound, domestic and outbound data in most cases. However, this should be decided on a case-by-case basis as MSISDNs have some advantages in some cases.

Because IMSIs change more often than MSISDNs (if the MSISDN changes, the IMSI will also most probably change; however, if the IMSI changes, the MSISDN might not change) for the life of the subscriber, the use of an MSISDN would produce better results for calculating the continuity of a subscriber (long-term visitors and repeating visitation) specially for domestic subscribers. However as MSISDN is sometimes not stored as the main identifier of the subscriber and because in outbound and inbound roaming cases IMSI is more reliable, the use of IMSI is suggested as the main identifier.

In case of inbound roaming and domestic data, on the BSS level the IMSI is translated to a TMSI (Temporary Mobile Subscriber Identity) for security reasons. A TMSI is randomly assigned by the MSC to every mobile in the location area (LA), the moment it is switched on. The number is unique within one LA, and it is updated each time the mobile moves to a new geographical area (new LA).

The network can also change the TMSI of the mobile at any time. And it normally does so, in order to prevent the subscriber from being identified, and tracked by eavesdroppers on the radio interface. This makes it difficult to identify mobile devices, except briefly, right when the mobile is switched on, or when data in the mobile becomes invalid for

one reason or another. If data from the BSS level is to be used in tourism statistics, IMSI-TMSI mapping should be conducted beforehand to provide a permanent single unique continuous identification code to a subscriber unless a TMSI is specially used for privacy reasons as a temporary unique code for a subscriber (see Section 4.2.5). However, this is discouraged as a TMSI might repeat over a period of time for different subscribers and thus cannot be considered a unique identity code over a longer period of time. On the NSS and NMS (central) levels, MNOs operate mostly with IMSIs.

4.1.5. Identifying the Country of Residence for Inbound Events

For inbound roaming, a procedure of identifying the country of origin of the inbound roaming subscriber is necessary. It is rather easy to extract the country code (MCC – Mobile Country Code) from an IMSI or CC from an MSISDN. In case of aggregation or other type of anonymisation/tokenisation, the process of identifying the country of origin has to be made beforehand. Examples of the IMSI and MSISDN identifiers are given in Table 3 and Table 4.

An IMSI consists of:

- MCC – Mobile Country Code;
- MNC – Mobile Network Code;
- MSIN – Mobile Subscription Identification Number.

Table 3. Example of IMSI representation. (Source: http://en.wikipedia.org/wiki/International_mobile_subscriber_identity).

IMSI:404685505601234			IMSI: 310150123456789			IMSI: 460001234567890		
MCC	404	India	MCC	310	USA	MCC	460	CHINA
MNC	68	MTNL delhi	MNC	150	AT&T Mobility	MNC	00	CMCC
MSIN	5505601234		MSIN	123456789		MSIN	1234567890	

An MSISDN consists of:

- CC – Country Code;
- NDC/NPA – National Destination Code/Number Planning Area (code of the MNO; this often means nothing, because it is a part of mobile number portability);
- SN – Subscriber's number.

Table 4. Example of MSISDN representation. (Source: http://en.wikipedia.org/wiki/MSISDN)

MSISDN: +380561234567

CC	380	Ukraine
NDC	56	Dnipropetrovsk
SN	1234567	Subscriber's number

A recognised country coding standard (e.g. ISO 3166) should preferably be used to map the countries of origin based on MCC or CC codes. Identifying the country of usual residence is closely connected to identification of usual residence and usual environment. The concept of identifying the usual environment for the purposes of identifying trips outside, is covered in Report 3a (Sections 2.2 to 2.4).

4.1.6. Time of the Event

Events are usually stored in MNOs' systems with the precision of a second. The time attribute should take into account the time zone and daylight savings time (DST – summer time) applicability of the event and should be transformed to the local main time zone of a country when needed (or a single universal main time zone in case of large countries with multiple time zones). All events in different MNOs should use the same time zone so there are no misunderstandings for obvious reasons. This is especially necessary for outbound roaming events as it is not always clear if the event time zone is stored according to the country where the event took place or the time zone of the home MNO.

4.1.7. Location of the Event (Geographical Referencing)

In case of inbound roaming and domestic datasets, geographical referencing means providing the most accurate location attribute to each event. The stored event record itself does not include any coordinates, but instead should have (not always though) the reference identity of the active antenna (and sometimes more signalling attributes). CGI (Cell Global Identity) is the standard code for the antenna globally. Although CGI is a specific standard, MNOs tend to create use their own antenna identity codes. Therefore, from here on, the identity of the antenna is named – *cell_id*. This *cell_id* can be linked to a geographical place through a database of antennae (usually in the OSS – Operation and Support Subsystem) that has a geographical attribute assigned to each antenna. There are two options for MNOs to provide coordinates for the events:

1. Assign coordinates using their own internal (secret) logic and processes and provide events with geographical coordinates;

2. Provide an antennae reference table in addition to events data for geographical referencing in the further processing.

In the first case, no further geographic processing is required (given that the referencing is accurate and adequate). The second case requires geographical referencing to each event. In this case there are four options to retrieve geographical attributes depending on the technical configuration of the MNO:

a) Single-point coordinates for each event via a cell database (Figure 6);
b) Single-point coordinates with some physical parameters about the cell (type, direction, power, antenna height, etc.) and about the event (timing advance from the antenna, signal strength, connection properties, etc.) that would help improve the spatial accuracy of each event;
c) Polygon coordinates of the cell coverage area;
d) Polygon coordinates of the cell coverage area (network plan) with physical parameters about the event (timing advance from the antenna, signal strength and connection properties).

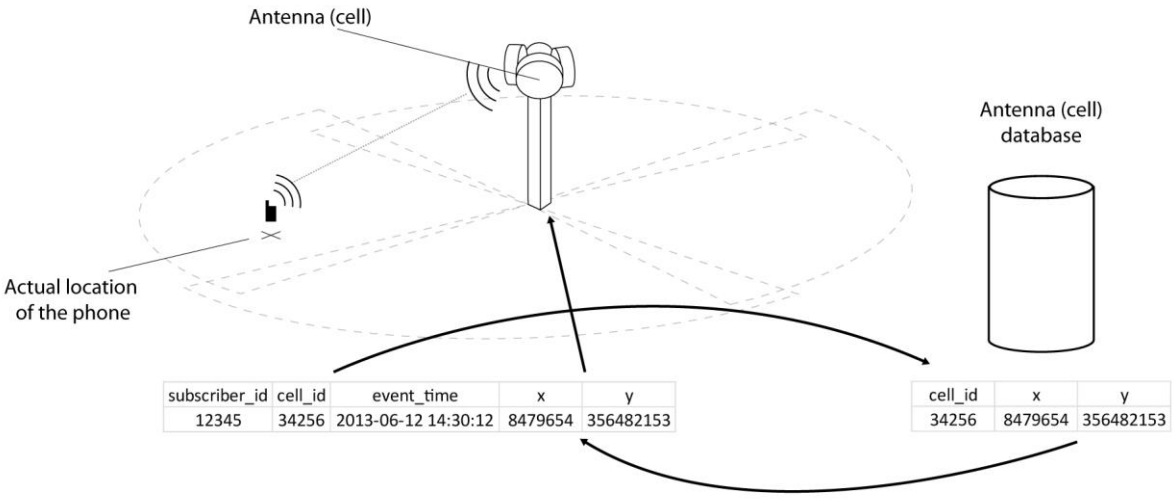

Figure 6. Illustration of the event coordinate mapping (single point mast coordinates).

Option A is the easiest option for almost all MNOs, but is also geographically least accurate. However, for more accurate spatial analysis, a probabilistic distribution of events in space should be conducted.

In case of outbound roaming events, the UDRs from roaming partner MNOs may occasionally include the CGI of the events that would provide more detailed geographical location attributes. However, it is usually not included as it has little value for local MNOs.

Therefore, a foreign country is by itself a geographical attribute for the outbound roaming data for each event.

It is important that geographical reference data is updated regularly as the network infrastructure changes rather often within MNOs – new antennae are added, some antennae change directions or are replaced. Geographical referencing is further discussed in Report 3a (in Section 2.1.2.).

4.1.8. Additional Attributes

There might be a number of different additional attributes of which some may be used to improve the quality of the methodology used. However, it is difficult to standardise them for use in tourism statistics, as some MNOs might be able to extract them and some might not. The additional attributes can be derived for events data, for antennae data and for subscribers.

4.1.8.1. Events Data Additional Attributes

Events data from MNOs can include various additional attributes besides the necessary attributes. These attributes can provide a more in-depth description of the event:

- type of the event (e.g. incoming call, outgoing SMS, location update, etc.);
- technical description of the event (e.g. duration of the call, technical parameters of the quality of the call, receiving/sending subscribers' identities);
- technical parameters of the phone and network communication (time advance from the cell, network standard used during the event (2G, 3G, 4G).

4.1.8.2. Network Data Additional Attributes

Additional parameters for antennae can be used to improve the accuracy of the events (see 4.1.7). Depending on the available attributes about the network, the appropriate location calculation of the event has to be used. Additional attributes for the network data can be type, direction, power output, height, coverage area of the antenna, etc.

4.1.8.3. Subscribers' Additional Attributes

MNOs can possibly provide access to additional information about domestic and outbound subscribers. Such attributes can be:

- different socio-demographic characteristics of the subscriber (owner of the contract) that might be age, gender, preferred language, etc.;

- details on the contract and the service:
    - private or business client;
    - invoice address;
    - average cost of the service;
    - contract type (pre-paid, post-paid SIM, machine-to-machine SIM).

These attributes carry highly sensitive information, but would provide additional possibilities for classification of domestic tourists, e.g. by age groups, gender, elimination or classification of specific subscribers, e.g. eliminating M2M SIMs, distinction of pre-paid subscribers as potential tourists, etc. However, there are several problems with socio-demographic attributes as there are many times when the actual phone user differs from the person marked on the contract – parents buy mobile phones for their kids, businesses provide work phones for employees, etc. So there are always a number of subscribers whose information is not correct and therefore such information has to be used with caution.

4.2. Preparation for Further Processes

Before proceeding to frame formation and data compilation, an MNO has to prepare, format and check for the quality of the data. After the preparation, a dataset can either be provided for the processing party (e.g. NSI) outside of the MNO's infrastructure, or processed in-house depending on the regulatory restrictions or agreement with the data processor (see 4.3). The result of this preparation is a dataset that is ready to be used in the further frame formation and data compilation processes (described in Report 3a).

4.2.1. Removal of Non-Representative Data

Many mobile devices are used for machine-to-machine (M2M) services and often do not represent a real human (vehicle telematics and tracking, gate opening devices, vending machines, etc.). The proportion of such devices is constantly growing and can create biased results as they often do not belong to the target population. Such devices can mostly be removed (or classified as M2M subscribers for later elimination):

- by a reference from the MNOs customer database (for domestic and outbound data) as M2M SIMs;
- by filtering out subscribers with irregular event behaviour like no CDRs and only DDRs, the pattern of the events is too high or too abnormal and not likely to be created by a human (for inbound, domestic and outbound).

This process result should be limited (mostly) only to human mobility representation in the dataset.

4.2.2. Eliminating Black-list Subscribers' Data and Sampling

MNOs might have a so-called black-list of subscribers who, for some reasons (security, extra privacy protection layer, etc.) have to be excluded from the prepared dataset. A similar procedure is used when a random sample of subscribers is used for further processing as an extra privacy protection option. Both cases require specific methods to delete the events of the elimination list of subscribers from the prepared dataset.

4.2.3. Formatting the Event Data

Formatting the prepared data to the agreed specification and format is required in order to recognise the provided data. The sequence of the fields, delimiters, field formats, file naming convention, file types and other specifications have to be agreed upon and followed by both sides. An example of a simple event package (inbound roaming data) with two events representing the identity of the subscriber, event time (presented in UNIX epoch format: number of seconds passed from Jan 1st 1970), antenna ID and country code (ISO 3166-1 alpha-2 standard):

```
subscriber_id,event_time,cell_id,country_code
34958620397862,1365589660,8433246,FR
98234756927843,1365589665,7634039,UK
```

A processing error or serious calculation error would result if a specified format were changed without notice. The specific format should be agreed on by both sides.

4.2.4. Initial Data Quality Assurance (QA) of the Event Data

The preparation process has to involve basic quality assurance before the data is provided for further processing. This should eliminate the possibility of data-based errors during the process (see Figure 7). Quality assurance could be conducted at any time during the preparation process though the most logical time is after the data has been formatted and before delivery for further processing.

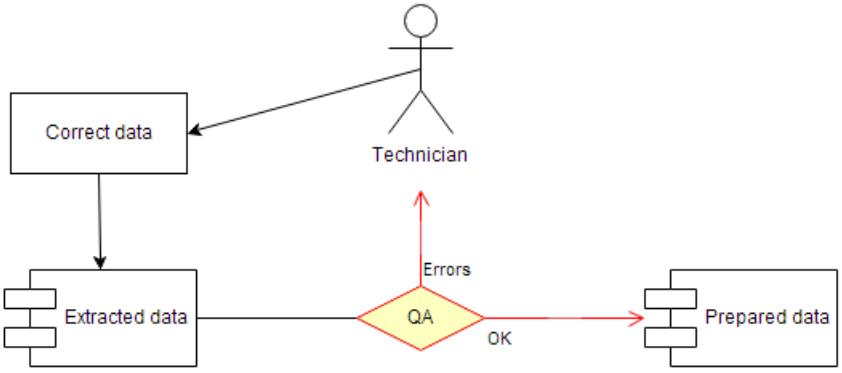

Figure 7. Simplified scheme of quality assurance procedures.

The QA process should check for the following errors:

1. Duplicate events in the dataset;
2. 'Gaps' in the data:
    - No data in the specific period of time;
    - Missing data for some antennae groups;
    - Missing data for specific subscribers groups;
3. Problems in mapping the countries:
    - New countries in inbound and outbound roaming (e.g. South Sudan);
    - Two or more home countries for inbound roaming subscribers;
    - No country codes for events in outbound roaming data;
    - No home country codes for inbound roaming subscribers;
4. Problems with network antennae/geographical references:
    - No geographical attributes in the antennae reference table;
    - False geographical coordinates in the antennae reference table;
    - Changes of antennae information not updated;
5. Event time logical and format control;
6. Additional attributes control (if such are provided);
7. Completeness of the data (whether all of the data for the specific period are provided);

If problems or errors occur during quality assurance, there are usually two options:

1. Corrections can be made:
    - immediately;
    - later (during the next data update). In this case:
        - data will be provided as are and corrections will be provided with the next data update;

- data will not be provided until corrections are made.

2. Corrections cannot be made for some reason (data permanently missing).

The specifics of the QA and error correction are a subject of a concrete MNO and its system. Obviously there should be a procedure that ensures the correction of the problems without long delays, effect on workflow (both the MNO and further processing) and the quality of the statistical results. A separate mechanism for feedback on flawed data should also be included if problems occur with the data during further tourism processing.

4.2.5. Handling Personally Identifiable Information

For event data, there are five options for handling the subscribers' identity code from the point of view of the privacy protection:

1. Personalised (not anonymous) permanent unique code – e.g. an IMSI is kept as a unique identity code of the subscriber (the identity of the subscriber is not protected further in the process if this process is conducted outside the premises MNO). If the prepared data is further processed outside the MNO, there is a possibility to combine the data from different MNOs based on this identifier. This can be considered as an 'ideal' initial data scenario from the point of view of the methodology.
2. Tokenised or hashed permanent pseudonymous unique code (the identity of the subscriber is poorly protected further in the process if this process is conducted outside the MNO) – this can be considered as an almost 'ideal' initial data scenario (identical to the previous if the tokenisation or hashing algorithms are the same in different MNOs and the codes can be combined with other MNOs' subscriber IDs).
3. Tokenised or hashed temporary unique code (the identity protection level depends on the time period of the code) – this method will reduce the quality of the initial data as the continuity of a single subscriber's identity is limited and thus not longitudinal. TMSI (see Section 4.1.4) might also be used for such token, however this is not recommended. is specially used for privacy reasons as a temporary The lifetime of the ID code can be 90 minutes (this option can be considered anonymous but is very limited in terms of methodology), 24 hours, 1 week, 1 month, 6 months, 2 years, etc.
4. Aggregated raw data – this can also be considered as fully anonymous data. Aggregation disables the possibility to calculate various important indicators in

tourism (and other domains). However, aggregated raw data can still provide a significant overview of tourism statistics and are still usable for some objectives. Aggregation of the raw data can be based on the number of subscribers within space unit in a period of time (e.g. number of subscribers per day per administrative unit). Annex 7 of the current report present the examples of different levels of aggregation that can be used for the illustrated artificial situation.

5. Other anonymisation algorithms that might or might not preserve longevity of the data. Such algorithms might involve false enrichment of the data (perturbation, obfuscation), time shifting, geographical shifting or other possibilities. These options will not be covered in this report as there is no existing and known anonymisation algorithm at this moment that would make the mobile positioning data fully anonymous (no direct or indirect identification possible) and preserve the longevity of the data.

See also Figure 9 for illustration of different variations of the provided data. Table 5 describes the limitations of the different options.

Table 5. The effect of different privacy protection scenarios compared to the 'ideal' scenario.

Anonymisation level of the subscriber's ID code	**Level of protection of personally identifiable information (anonymity)**	**Limitations compared to the 'ideal' scenario to some of the methodological aspects**
Permanent unique code (IMSI)	Not protected at all.	The ideal scenario. • ID combining with other MNOs: YES • Domestic activity space recognition: YES • Usual environment: YES • Trip identification: YES • Same-day/overnight visits: YES • Border bias: YES • Long-term visitors: YES • Transit visits: YES • Repeating visits: YES • Visit routes: YES
Permanent pseudonymous unique code	Very weakly protected. Pseudonymous, but not anonymous. Indirect identification is possible.	Almost ideal scenario. Some limitations if different MNOs use different tokenisation/hashing algorithms and subscribers from different MNOs cannot be combined. • ID combining with other MNOs: YES (if the same algorithm is used) / NO

		• Domestic activity space recognition: YES • Usual environment: YES • Trip identification: YES • Same-day/overnight visits: YES • Border bias: YES • Long-term visitors: YES • Transit visits: YES • Repeating visits: YES • Visit routes: YES
Temporary pseudonymous unique code 2 years	Very weakly protected. Indirect identification is possible.	Good scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: YES • Usual environment: YES • Trip identification: YES • Same-day/overnight visits: YES • Border bias: YES • Long-term visitors: YES • Transit visits: YES • Repeating visits: LIMITED • Visit routes: YES
Temporary pseudonymous unique code 6 months	Very weakly protected. Indirect identification is possible.	Good scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: YES • Usual environment: LIMITED (6 months) • Trip identification: YES • Same-day/overnight visits: YES • Border bias: YES • Long-term visitors: NO • Transit visits: YES • Repeating visits: LIMITED • Visit routes: YES
Temporary pseudonymous unique code 1 month	Very weakly protected. Indirect identification is possible.	Weak scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: YES • Usual environment: VERY LIMITED • Trip identification: YES • Same-day/overnight visits: YES • Border bias: YES • Long-term visitors: NO • Transit visits: YES

		• Repeating visits: NO • Visit routes: YES
Temporary pseudonymous unique code 1 week	Weakly protected. Indirect identification is mostly possible.	Weak scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: LIMITED • Usual environment: NO • Trip identification: LIMITED • Same-day/overnight visits: LIMITED • Border bias: YES • Long-term visitors: NO • Transit visits: YES • Repeating visits: NO • Visit routes: LIMITED
Temporary pseudonymous unique code 24 hours	Very well protected. Indirect identification is possible but in very few cases.	Very weak scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: NO • Usual environment: NO • Trip identification: NO • Same-day/overnight visits: LIMITED/NO • Border bias: LIMITED • Long-term visitors: NO • Transit visits: LIMITED • Repeating visits: NO • Visit routes: LIMITED
Temporary pseudonymous unique code 90 minutes	Excellent protection of the identity (can be considered fully anonymous). Indirect identification almost impossible.	Very weak scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: NO • Usual environment: NO • Trip identification: NO • Same-day/overnight visits: NO • Border bias: NO • Long-term visitors: NO • Transit visits: NO • Repeating visits: NO • Visit routes: NO
Aggregated raw data	Subscribers' identities totally protected if threshold is also used (e.g. not showing aggregates under 10).	Very weak scenario. • ID combining with other MNOs: NO • Domestic activity space recognition: NO • Usual environment: NO • Trip identification: NO

		• Same-day/overnight visits: NO • Border bias: NO • Long-term visitors: NO • Transit visits: NO • Repeating visits: NO • Visit routes: NO

An additional option to improve privacy protection is to use a continuous limited random sample of subscribers instead of 100% of all subscribers (census). This measure allows the assumption that it is not known if a specific subscriber is included in the sample or not, thus aggravating any potential direct or indirect identification of the subscriber from the population frame. The use of a sample will result in smaller amount of the data and might have an effect on the quality and the representativeness of the data. These aspects are discussed in the Report 3a (Sections 2.1.1.1. and 3.2.2.).

From the point of view of accessibility to the data, the use of sample can benefit in the resources needed for storing and processing the data (less data to process), but this is only beneficial in very large MNOs where the use of random sample results in acceptable sampling error. As presented in Report 3a, the smaller the domains are (smaller regions, time frames) the larger are discrepancies compared to the census.

If sample is used, MNOs have to implement one more additional component allowing sampling based on the subscribers. The sample should be taken not based on events, but on subscribers, i.e. use the percentage from the subscribers list and process all data of such sample, therefore just providing a percentage of the data does not achieve the objectives of useful base data.

4.2.6. Delivering the Event Data to the Processor

By the end of the initial data preparation, either event data or aggregated data is provided for further processing. In case of the event data, each event record should have the unique identity of the subscriber, the location attribute and time attribute. In case of aggregated data, further processing is rather limited.

Depending on the location of the further processing (see 4.3), the next phase of the process can be located within the MNO's infrastructure or outside on the premises of the statistics authority (see Figure 8). If the process continues within the MNO, the specifics of delivery to the processing system are a technical matter of the MNO. This is probably a

separate designated system that handles this processing. This system is responsible for final outputting of the result data to the end user (e.g. NSI).

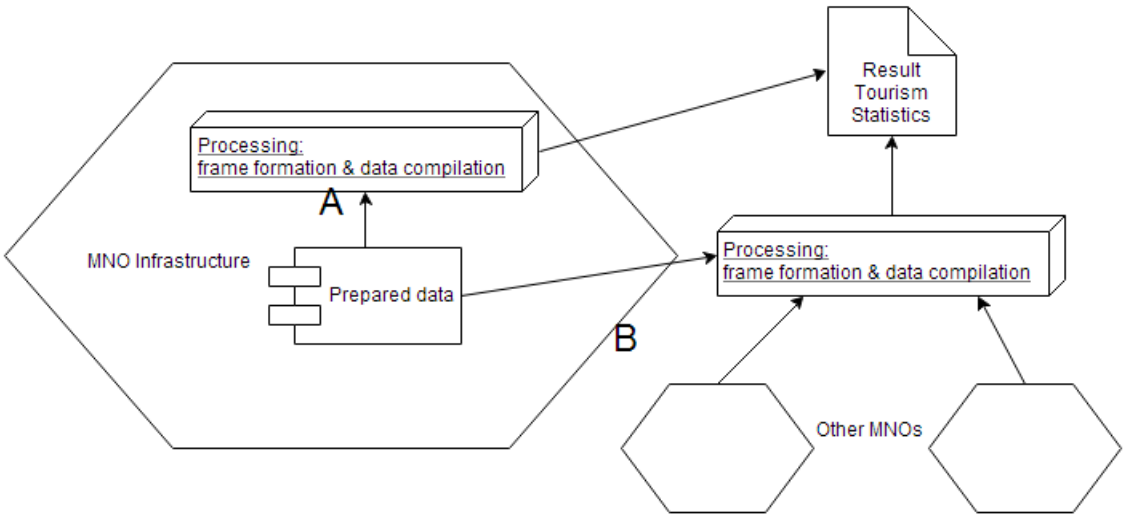

Figure 8. Two options for further processing the prepared data: internally within MNO (A) or externally (B).

If the further processing takes place outside the MNO's infrastructure, the method and security settings of the transfer of the data packages should be agreed on between the MNO and the processor. An important aspect is also the data update frequency – how often MNOs can/must provide the data (near-real time, daily, weekly, monthly, etc.).

4.3. Allocation of Data Processing Components

There are two main, advisable (from the aspect of methodology) options for obtaining data from MNOs according to a statutory obligation to collect and otherwise process the mobile positioning data:

a) extraction, preparation and transmission of un-processed data to the statistics authority who will conduct the processing (frame formation, data compilation, 1 and following options from Figure 9);
b) extraction, preparation and processing the data within the premises of MNOs and transmitting the aggregated results to the statistics authority (2 and following options from Figure 9).

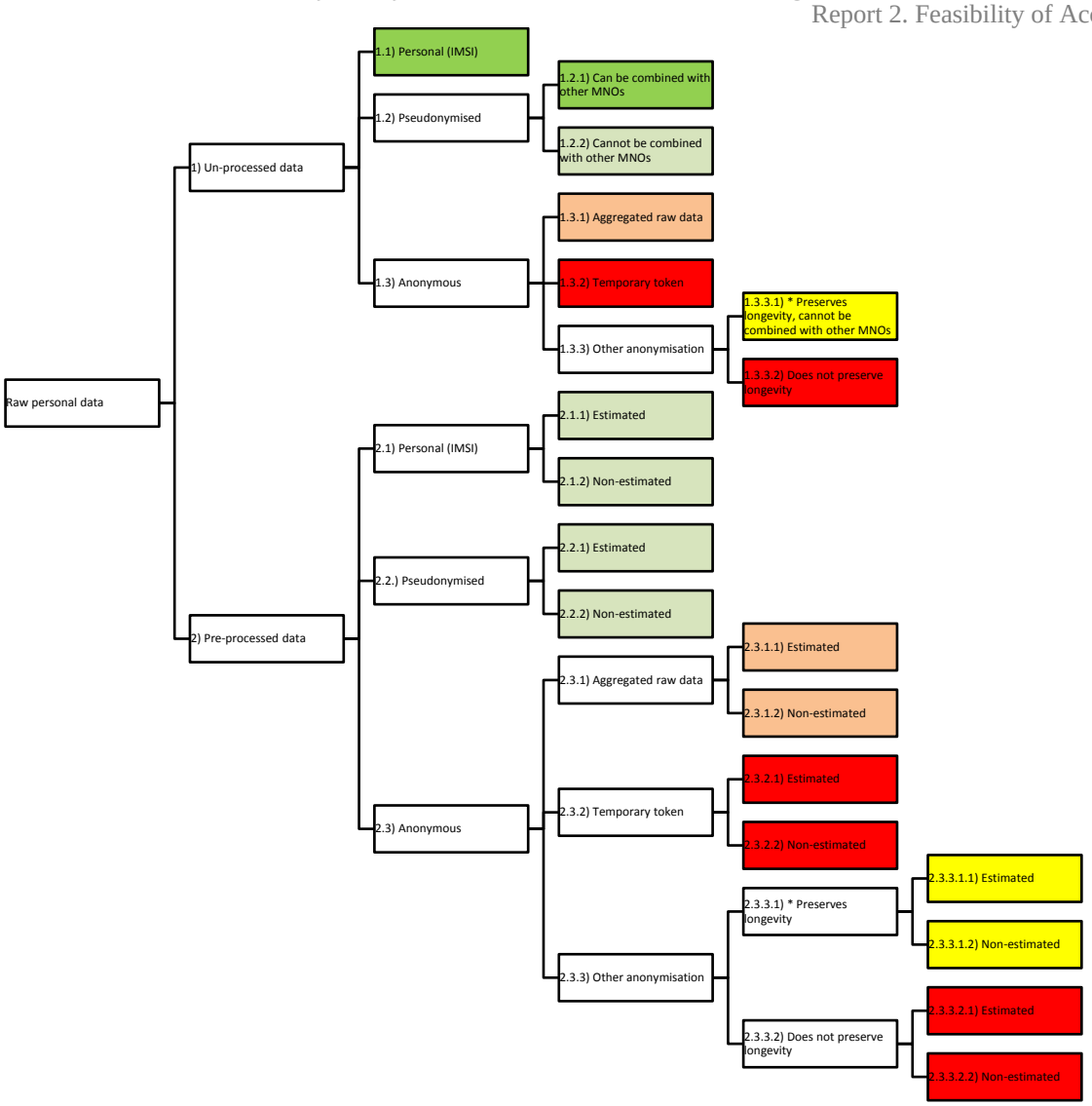

Figure 9. Options of how the data can be delivered to the NSI based on the level of anonymisation of personally identifiable information of the data from MNO raw databases (left) to the point where data is provided to the NSI (right ending of each branch). The colours represent the relative level of usability in terms of quality of the end results. Green representing the best option, red representing the worst. * - currently not known if such algorithms exist that can fully anonymise the dataset (directly and indirectly unidentifiable data) and at the same time preserve the longevity of the data for single subscribers.

If there is no statutory obligation for MNOs to provide the data and statistics authorities to process the data, but for some reasons (research, secondary supporting data) the data transmission is established (voluntary basis, commercial agreement), there are several options for transmitting the data. In this case most probably MNOs do not have the legal right to transmit directly (1.1 from Figure 9) or indirectly (1.2 from Figure 9) identifiable (personal) data to NSI. Other options are a subject of local legislation and can be very limited (e.g. 2.3.1.2 and 2.3.2.2 from Figure 9 in German and French example) or somewhat useful (1.2.2, 2.2.1, 2.2.2 from Figure 9 in Estonian example).

The 'ideal' scenario represents the options 1.1, 1.2, 1.3.3.1, 2.1, 2.2, 2.3.3.1 from Figure 9. There are several options for allocating the specific technical components responsible for different sections of the whole data processing chain. Those components can be distributed between the MNOs (internal) and external receiving party (e.g. NSI). Four examples of the distribution of the processes for 'ideal' scenario are presented in Figure 10 and depend on the location of individual processing system parts. This is an indicative suggestion and the actual allocation depends on the decision and agreement in the Member States individually.

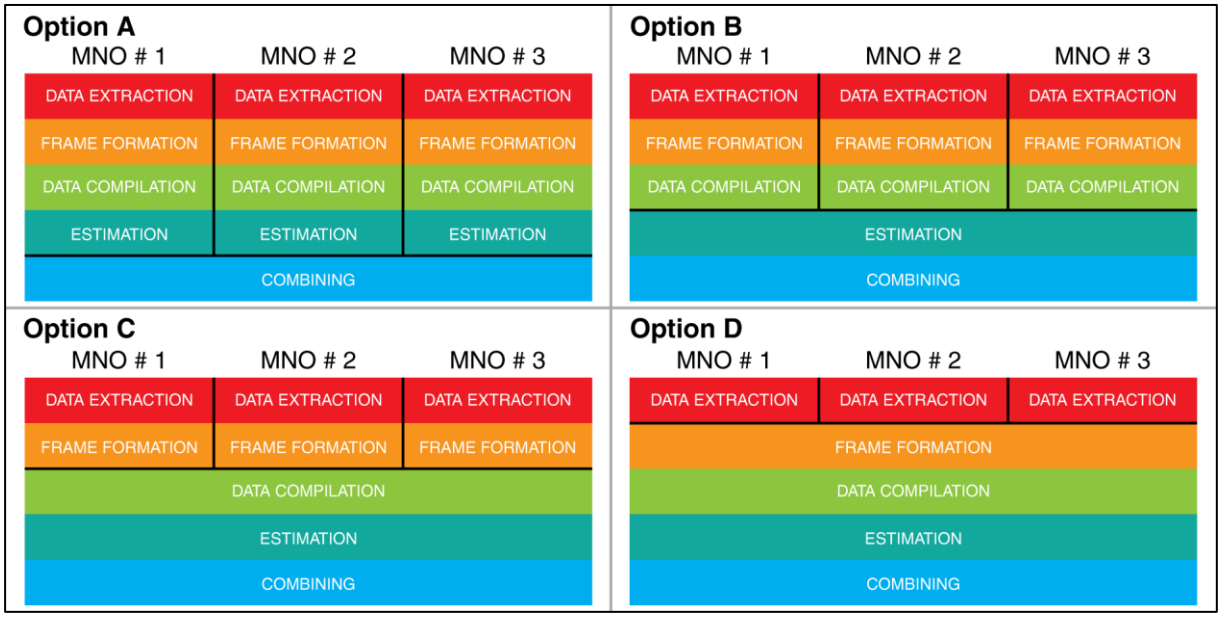

Figure 10. Four options of distribution of processes between MNOs' and external data receiving party (e.g. NSI).

From the overall cost point of view, option A is the most expensive as all MNOs have to implement the systems for processing the data up to the aggregated results (biggest burden). The implementation costs on the side of the receiving party are small as only the system for combining the results is required. Option D presents the smallest burden on the MNOs, however the cost of the system implementation on the receiving party is rather high. Still this is the least expensive option from the point of view of the total costs. The elaborated discussion and example of the cost calculation is presented in Section 5.1 and in Report 4.

From the point of view of the methodology, option D is the best as the responsibility of the whole processing lies on the receiving party and MNOs only have to extract and provide the initial event data. Option D also provides the possibility to combine same devices that use different MNOs' roaming services (inbound data). Otherwise MNOs have to internally take into consideration the corresponding coverage issue (see Section 3.2.1. in Report 3a).

4.4. Differences in Network Systems

There are several network equipment and system providers for MNOs out of which Huawei, ZTE, Ericsson, Nokia Siemens and Alcatel-Lucent, Motorola are the largest providing the whole technological chain. There are a number of smaller vendors providing some specific system parts. All equipment providers have to follow local (state, EU) or international standards on basic equipment. A standard example of basic network structure is illustrated in Figure 5. The variation in basic equipment is usually expressed in extra features, some of which can be relevant to the storage and representation of location data. For example some vendors provide extra layers for storing all location events from the network (not used by most MNOs by default) for security applications, business intelligence applications, marketing or network management purposes. However, such applications are not standardised and are usually custom-developed and implemented according to the specific requirements of the MNOs.

Mostly the differences within the networks relevant to this study lie in the settings, the specifications and processes of the data flow. For example the amount of data stored in different registries and databases, the length of the stored data (important from the regulative point of view, data retention requirements – see 3.1.1.3, Directive 2006/24/EC), the attributes of the events that are stored, the update intervals for the data to be relayed to specific processes, etc. Such settings depend on how much analysis the MNO is conducting based on that data.

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. Such data might be for example the CDRs in a billing database or a data warehouse. Any extra data usually requires additional technical implementation.

Based on the findings during the study, the basic minimum dataset that can be used for tourism statistics does not depend on the network equipment provider as the initially usable data falls under the data retention regulation. The differences express themselves in the quantity of additional data that is stored due to extra hardware/software modules provided to the MNOs.

4.5. Patents and Intellectual Property Rights

The relevance of understanding patents and intellectual property (IP) rights for this feasibility study is that it provides an understanding of on-going technological developments

and their direction as they take place in the European and global market place. The collection and processing of data for tourism statistical purposes – as explained in previous chapters – is technology intensive and sees the application of a range of data processing methodologies and techniques.

With an in-depth understanding of patents and IP rights potential opportunities and constraints may be revealed for the application of data collection, storage, extraction technologies and processing methodologies for statistical analysis i.e. as to whether some are subject to restrictions in use and barriers, including mainly financial and economic as a result of patent and intellectual property right holders influence. Having this overview thus will result in short explanation of the intellectual property rights systems and potential patents that might affect the licencing of the relevant technology.

Because the actual relevance and potential infringement threat is fairly difficult to establish, it is not within the scope of the current report to declare the patents that act as barriers in generating tourism statistics. The presented list is merely an indication of the patents that might or might not be relevant to the technology used to generate tourism statistics. However such list might also present opportunities to involve technologies mentioned as their use might improve the quality of the data or simplify some processes. A search result for potentially relevant patents in EU and the World is presented in the Annex 6.

4.5.1. Introduction to Patents

Patents are intellectual property rights (hereinafter ‘patent’), a public title of industrial property that gives its owner the exclusive right to use his/her invention in the technical field for a limited number of years. A patent needs to be applied for and should describe an invention, i.e. a new solution to a technical problem which satisfies the criteria of being novel and it must involve a non-obvious inventive step. A patent may be granted to a firm, an individual or a public body by a patent office. It remains valid in a given country or area for a limited period of time as designated in the application.

The patents can be categorised as EU patents and world patents. The European patent framework is governed by the European Patent Convention (EPC) that establishes a uniform patenting system for all countries signatory to the Convention; the countries in scope for the Eurostat feasibility study – Estonia, Finland, Germany and France – are all signatories. A granted European patent is protected under national law in each of the countries designated in the application.

At the central level the European Patent Office (EPO, based in Munich) is the authority that grants European patents. These patents can be referenced in the European Patent Register database, which is managed by EPO.

World or international patents are filed under the Patent Cooperation Treaty (PCT) allowing a single initial application to cover one or more of the 148 contracting states. After a certain period of time, the PCT application is converted into individual applications in each country where protection is needed.

At the global level there is no centralised patenting organisation, as the EPO for the European region, to handle filing of patents, searching and examination. Under the PCT there are a select International Searching Authorities (ISAs) and International Preliminary Examination Authorities (IPEAs), where initial filing, searching and examination is performed, prior to referral to the individual country patent offices for further processing. The EPO belongs to the group of ISA/IPEA and can therefore receive international PCT applications.

Furthermore at the global level there is the World Intellectual Property Organisation (WIPO) that promotes and facilitates cooperation and data exchange between individual country patenting and IP offices. As part of this it offers the Patentscope Database for referencing international PCTs.

The most prominent classification schemas, under which patents are (and till recently have been) filed, are the European Classification (ECLA) administered by EPO, and the US Patent Classification (USPC) administered by the US Patent Office (USPO). The USPO and EPO have initiated a harmonisation of classification schemas, which has resulted in the Cooperative Patent Classification (CPC). The EPO has completely switched to the CPC, and the USPO is expected to do so by 2015. Other countries are also following suit with the Korean patenting authorities piloting the use of CPC.

At the global level there is still the International Patent Classification (IPC) administered by the WIPO. Both IPC and CPC schemas follow one another closely in structure.

On European level patents are granted but not enforced by the EPO. The EPO's main task is to receive patents, examine and register/publish these with country/jurisdiction designation. As such European patents are enforced at the national country-by-country level in line with national patent laws (which are ideally but not necessarily harmonised at EU level).

Infringements in the most basic terms occur if within a jurisdiction patented rights are used without the patent holder permission. There are no actual oversight bodies or organisations that keep track of all products developed within society and whether patents are infringed or not. Identification of patent infringement is mostly done by patent holder and is based on their proactive observing and defending of their position in the market place. Undoubtedly if a patent holder has gone through the efforts to registering a patent, they are likely to also keep a watchful eye.

At the submission of patent claims by prospective patent holders, the extent of protection is determined at the technical and substantive level by claim descriptions, technical drawings etc. all related to the product, system or method to be patented. In addition the extent of protection is defined by designation of jurisdictions i.e. EU countries where protection is granted.

In the case where there are patent infringements of a European patent in more than one EU countries, still litigation needs to take place at the national courts in the individual countries (Appelt, 2007). Cross-border injunctions have been limited by the European Court of Justice (GAT, 2006; Roche Nederland, 2006) that argues that European patents are national rights, and therefore must be enforced nationally.

Avoiding infringement is considered beneficial as it can for example lower/or prevent future legal costs in resolving patent infringement litigation cases or potential injunctions by the patent holder. As such product, systems, or methodologies developers can take initial steps to avoid infringement (or lower the risks of) by conducting among others online patent research, review products from competitors. In addition formal screenings can also be conducted with specialised assistance e.g. patent lawyer or similar services provider.

Based on the above, if patent infringement is determined as likely, permission can be asked from the patent holder to use the patent under license. Licensing of patents is a very varied legal area with many different types of licensing agreements. This section will not cover the licensing in-depth.

Utility models provide an alternative to the patent model in which protection maybe obtained an invention. Like a granted patent, a utility model is an exclusive right granted for an invention, which allows the right holder to prevent others from commercially using the protected invention. Like a granted patent, a utility model is an exclusive right granted for an invention, which allows the right holder to prevent others from commercially using the protected invention. In its basic definition a utility model is very similar to a patent, except

that the requirements for acquiring a utility model are less stringent than for patents. In practice, protection for utility models is often sought for innovations of a rather incremental character, which may not meet the patentability criteria.

Relevant for this feasibility study is the fact that utility models can be used 'strategically' by service providers that are involved in extraction and processing of mobile positioning data, should relevant systems and methods as presented in patent applications face challenges in the examination stages (e.g. with the EPO). As such utility models can be used in other situations and perhaps from a more strategic point of view in order to enhance a company's IP portfolio.

4.5.2. Summary of Relevant Patents

Relevant patents are listed as an Annex 6 to this report. It should be noted that the listing is an initial selection of patents. In the haystack of thousands of patents there possibly exist additional patents which can be identified depending on the variation and scope of search variables.

Insight in the patents relevant for tourism statistics provides a greater knowledge on available methods and systems for data collection and processing. Clearly, most patents are developed for the US market place, with very few making it into the European marketplace with patent granted status. This opens opportunities for European market entrants into the domain of tourism statistics generation in the sense that they can learn from the methods and systems being used elsewhere, and try to apply them (albeit with some slight modifications) in the European market.

Insight in patents further reveals that patents for capturing and registering location data from mobile terminals are plenty; less are there patents for extraction and use, and in particular linking to demographic datasets. There are however some patents that try to capture the end-to-end processing chain from data capturing to demographics based analytics, which are relevant for tourism market segmentation.

There are also few traffic management or traffic modelling methods and systems under patent status that can be valuable for tourism destination management planning, in particular from the perspective of facilitating good access and developing available (basic) services. As a last point, for market segmentation purposes, more demographic-linked tourism statistics are needed, of which there are almost no patents.

4.6. Continuity of Data Access

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

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

4.6.1. Changes of Characteristics of the Data

There might be several causes for the characteristics of the data to be changed. The reasons are mostly technological and are usually positive (benefiting the quality of the data and statistical results). For example, more events data can be stored in the future. In the EU the European Commission roaming cost regulations are directly linked to the increase of the number of inbound and outbound roaming events as it is cheaper to use the phone while on the trip. The increase of the number of events (per subscriber) would improve the quality of the data and produce more adequate statistical results (smaller coverage issues). But this might also cause shift in the time series thus affecting comparability over time.

The negative effects can be seen as the inability to exclude the increasing number of machine-to-machine (M2M) SIMs and the confusion caused by the number of SIMs owned and carried by one person. This topic is more thoroughly discussed in Report 3a (Section 3.2.1.).

The following Figure 11 summarises how the number of subscriptions has changed in the example countries of this study, and Figure 12 shows the number of subscriptions for European countries (EU, EFTA and Candidate Countries) at the beginning of the century and in 2012. The ratio of mobile subscriptions to the general population and the change thereof can also have an effect.

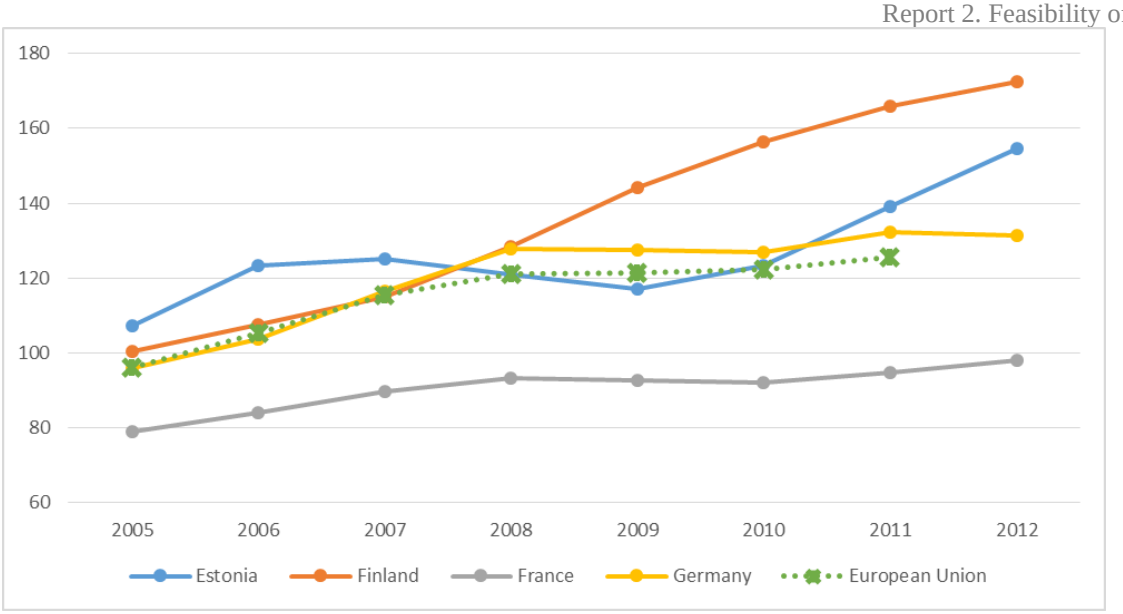

Figure 11. Change in the number of mobile cellular subscriptions (per 100 people) in Estonia, Finland, France, Germany and in EU from 2005 to 2012. Source: World Bank

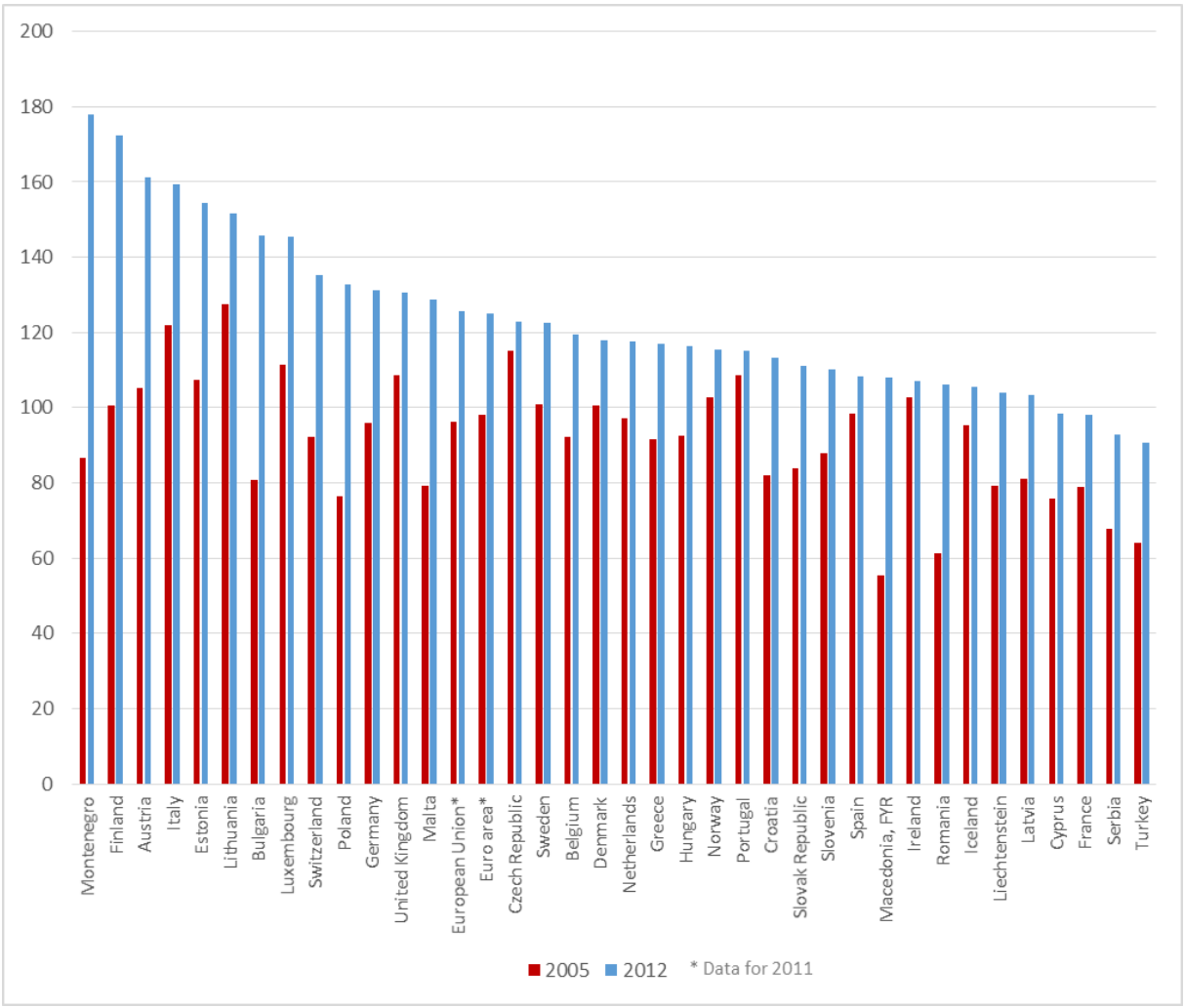

Figure 12 Mobile cellular subscriptions (per 100 people) in European countries (EU, EFTA and Candidate Countries) in 2005 and 2012. Source: World Bank.

Another 'threat' is the change in cellular technological standards. Currently the new 4th generation (4G) of mobile network infrastructure is being deployed in many countries, which allows using more bandwidth than the 3rd generation (3G). The 5th generation of mobile phones will most likely advance aspects of energy efficiency, sustainability and affordability of the use of mobile networks, which should not change the methodological aspects of the statistics from mobile networks. Those new standards do not potentially have negative effects on the current methodology.

4.6.2. Change in the Number of Data Providing MNOs

Change in the number of MNOs that provide the data might also affect the characteristics of the data as well as the methodology and processing limitations and rules. Such changes might also bring positive effects (more MNOs that provide the data results in the improvement of the quality of the data) but might also introduce a break in the time series. A change in the number of MNOs providing initial data would result in a change in the representativeness of the data. More MNOs means larger sampling frame and improvement in the representativeness and the quality of the results; however, this would also require corrections in estimates.

The quality of the data of each MNO providing the data should be quantitatively assessed. For example for inbound roaming data, each MNO should be analysed with regard to the distribution of foreign visitors according to their country of residence. This information can be used in the estimation to reduce bias due to the coverage errors. In ideal case, no corrections are needed and combining the data from several MNOs only improves the reliability of the results and does not affect the results themselves. In real life different MNOs have different penetration rates among different nationalities (see Figure 12).

The same logic applies to domestic and outbound data – results from different MNOs should be weighted to take into account the differences in the backgrounds of the subscribers. For domestic data, MNO's subscribers differ in their socio-demographic characteristics, geographically etc. Weights used in the estimator indicate the influence of each MNO and the assessment of the impact of the withdrawal of the MNO can be made. If a small MNO withdraws, the reliability of the data will decrease, but the effect might be rather insignificant. If an MNO with major importance among one subscriber group withdraws (e.g. large majority of the roaming subscribers from the major tourism donor country), this would cause biased results for this specific group.

4.6.3. Major Shift in Mobile Technology

A major effect on the methodology would be a drastic change in cellular technology and the principles of telecommunications. For example such a change would be the transformation of communications from ground-based networks to satellite communications or a change of mobile communications from mobile operators to some other businesses (IP-call providers like Skype, or a major technological revolution). However, the future role of MNOs is often discussed (Accenture 2013, ATKearney 2012, Cisco 2013, KPMG 2013) and most of the future scenarios can be considered suitable from the aspect of this study – MNOs are to remain technical providers of communications infrastructure. There will still be technological networks connecting to mobile devices that are used to mediate content services (voice, messaging and internet). This means there will still be a basis for logging the activities (events) with location reference and the attributes of the events. If this is true, the concept of location events can be used as a basis for mobility statistics calculations in at least the near foreseeable future.

4.6.4. Flexibility of Access, Flexibility to Introduce Changes to Methodology, Risk Assessment

Risk assessment of the potential threats to the continuity of the data based on abovementioned causes is presented in Table 6.

Table 6. Risk matrix of potential threats of continuity of the data. Probability: likely, moderate, unlikely, remote.

Threat	Probability	Impact	Possible action
Changes in the characteristics of the data (better data)	Likely (probably every few years)	Will affect the results if the new data decreases some of the coverage issues compared to the previous data. Positive effect.	Adjustment of the methodology, analysis of the coverage issues with new data, adjustment in estimation basis (smaller correction coefficients). The additional data can sometimes be ignored if the cost of the implementation of methodological changes overweighs the increase in the potential quality of the results.
Changes in the characteristics of the data (worse data)	Unlikely	Will affect the results as probably the coverage issues increase. Negative effect.	Adjustment of the methodology, analysis of the coverage issues with new data, adjustment in estimation basis (higher correction coefficients).

Changes in the characteristics of the data (substitution of previously used data e.g. DDR instead of CDR)	Moderate	Will affect the results as the coverage issues will be different. Unknown effect.	Adjustment of the methodology, analysis of the coverage issues with new data, adjustment in estimation basis.
Drastic changes in MNO market shares	Moderate (caused mostly by price wars, e.g. penetration pricing)	Will affect estimation coefficients. Moderate effect.	Requires analysis of the coverage issues, adjustment in estimation basis for all MNOs, adjustments in the combination process of MNOs. The process is similar to regular estimation adjustments (described in Report 3a).
New data providing MNO (existed previously, but was not providing the data)	Likely (if previously only some MNOs were involved)	Will affect the results as the data will be more realistic, the new estimations will be more precise, decrease in some coverage issues. Positive effect.	Analysis of the coverage issues, adjustment in estimation basis for new MNO, adjustments in the combination process of MNOs.
New data providing MNO (did not exist previously)	Unlikely (MNO market is rather established and changes do not happen very often)	Will affect the results as the market share of the MNOs might drastically change. Creates costs related to new data extraction and processing setup Moderate effect.	Requires analysis of the coverage issues, adjustment in estimation basis for all MNOs, adjustments in the combination process of MNOs. The process is similar to regular estimation adjustments (described in Report 3a).
New MNO in the country (does not provide the data)	Unlikely (MNO market is rather established and changes do not happen very often)	Will affect the results as the market share of the MNOs might drastically change. Negative effect.	Analysis of the coverage issues, adjustment in estimation basis for all MNOs.
Merging of two MNOs in the country	Moderate (Relates to	Might have some effect. However, the	Change in the systems of MNOs (combining previously two separate

(providing data)	cheap debt equity, relaxed EC decisions on EU anti-trust rules for MNO mergers, increasing popularity of shared infrastructure. However, merger is a slow process and cross-border acquisitions more likely)	amount of data does not change. Moderate effect.	data extraction and processing systems). Analysis of the coverage issues based on merging MNOs, adjustments in the combination process of MNOs.
Decrease in the number of providing MNOs (one decides to stop providing the data)	Unlikely (Relates to very drastic issues of privacy concerns or change of ownership of the MNO)	Will affect the results as the data will be less realistic, the new combined estimations will be less precise, increase in some coverage issues. Negative effect.	Assessment of the MNOs and the impact of losing one data source. If the data from a single MNO discontinues but there is a historical data that can be compared to the remaining MNOs, then the historical impact on the final result from one MNO should be mathematically calculated in order to see if there is a constant effect of this one MNO and if possible, continue applying the mathematical rule instead of the actual data from MNO. Sampling might be a solution in case of outbound and domestic tourism.
Decrease in the number of providing MNOs (one MNO seizes to exist)	Unlikely (In the case of MNO bankruptcies assets are likely acquired or restructured	Might have an effect on the results as the data will be less realistic, the new estimations will be less precise, decrease in some coverage issues.	Analysis of the coverage issues, adjustment in estimation basis for new MNO, adjustments in the combination process of MNOs. Same as above.

	instead)	Positive effect.	
Major shift in mobile technology	Remote (very difficult to foresee)	Will have a major effect on the results. Unknown effect.	Assessment if the methodology can be used with new data.

It is difficult to foresee the final effect of the changes on the results when the changes have just been made. In some cases changes might not affect the outcomes at all if the methodology and the estimations are adjusted properly. Ideally including all MNOs in the country should have insignificant increase in the quality compared to a single MNO with good coverage and adequately implemented methodology and estimations. Each change should be well assessed before any adjustment is made.

It is possible to measure the effect of each MNO on the combined resulting data. If such effect can be defined by some mathematical rule (e.g. correction coefficient of specific MNO), then in case of the losing one MNO, such rule should be used to keep the effect of the lost MNO. However usually such rule is very difficult to establish and if the MNO actually seizes to exist (bankruptcy), then other MNOs collect the subscribers and estimations have to be corrected. In such a case, the final outcome should be the same as previous.

If it is not possible to define mathematical influence of the lost MNO, the estimates for the remaining MNOs should be reviewed. However in ideal case, the estimates have to be close to the realistic number of tourists anyways.

If all of providing MNOs seize to provide the data, then there is little to do as historical data is not projectable to the future. Therefore the legal framework for using the data should minimize the possibilities for MNOs to bail out.

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

The flexibility to introduce the changes depends on the configuration and the setup of the system. From the point of view of the allocation of the technology (see Section 4.3) the changes are more easily adopted when the processing resides outside of the MNOs and MNOs only extract, prepare and transfer the data to the processing party (e.g. NSI). In case of new available data (e.g. DDRs included in addition to CDRs) MNOs only need to change the

extraction process to include the new data types. Altering this process is fairly simple. However the following processing might require more extensive modification. If the following processes are allocated in NSI, then it is easier to modify as the changes have to be made in one system. But if the processes are located in MNOs, then all MNOs have to update the system to be able to handle the new data and this is more costly and time consuming. Obviously all such changes require a period of testing when new and old system might be required to run in parallel.

5. Financial and Business-related Opportunities and Barriers

Based on the previous experience of consortium members and contacts with MNOs there are several obstacles for the MNOs to providing the data for usage outside the main business purpose. There are also financial aspects that need consideration. MNOs are business units and their direct objectives are to provide services to their customers and produce profit for owners. Still, many MNOs contacted within this study have shown interest in using the data outside their main business profile in order to either increase revenues, provide support for their main business, use the data for internal analysis and sometimes to support innovation (big data) and conduct projects within their social responsibility units to participate in developing community (transportation, urban planning, etc.). Practical aspects (negative and positive) that affect the decisions of MNOs to utilise their data are following:

a) EU and national-level legislation – the aspects of limitations and obligation for them to provide the data (mentioned by all respondents in the survey and interviews);
b) Privacy protection questions, effects of public opinion and possible decline in the number of clients due to bad reputation (mentioned by most of the respondents);
c) Preservation of business secrets in order to maintain business competitiveness compared to competitors (mentioned in some interviews with MNOs);
d) Cost of technological implementations and workload (resources) required to provide the data (burden mentioned by most of the MNOs who responded). Financial aspect – the cost of the data was also mentioned by several respondents on the user side;

e) New revenue possibilities through commercialisation of the data (mentioned in interviews by most of the MNOs as the reason for investigating the possibilities of utilizing their data);
f) Internal benefits and new insights from the analysis of the data in new manners (mentioned by some MNOs);
g) Participation through social responsibility, e.g. supporting analysis on regional development (mentioned in some of the responses);
h) The support to their main business, e.g. positive image towards the government (mentioned by some MNOs in the interviews).

The legislation and privacy protection issues are covered in Section 3 the current report. The current chapters concentrate on the other aspects of accessibility mainly from the MNOs' point of view based on experience and feedback from MNOs.

5.1. Implementation and Maintenance Cost of MNOs (Burden)

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

The cost of the system that retrieves and processes the data in MNOs consists of the implementation (initial investment required to set up the system) and maintenance costs (price for keeping the system working); and depend on several variables. These variables are:

- Size of the MNO / amount of data chunk to process;
- Allocation of the processing resources (i.e. whether MNOs only have to extract and deliver the raw initial data to the processing party outside or MNOs have to implement the full processing chain);
- The number of processes to conduct by MNO (i.e. forms of data: inbound and / or domestic and / or outbound; geographical probability calculation for usual environment, etc.);
- Maximum allowed latency (i.e. the maximum allowed processing time from the extraction of the initial raw data to the delivery of the data to NSI);

- External variables not foreseeable in this report (e.g. licencing of external technology, outsourcing costs and cost of internal resources – man-hours).

This chapter will propose two extreme opposite scenarios of the costs for implementing and maintaining the system for providing data for tourism statistics based on the allocation of parts of the system (see Option A and D – derived from Section 4.3, Figure 10). Costs presented here are rough indications only and should not be taken as facts because only MNOs themselves can provide the actual costs based on their internal calculations. The cost scenarios presented here cover the costs of a full processing chain (maximum extent) and only extraction and preparation of the data (minimum extent). The maximum and minimum extents are presented from the point of view of the MNO – maximum extent also means maximum burden and cost to implement and maintain the system. The latter basically means that the core of the processes is shifted from the MNO to the receiving party (e.g. NSI).

The calculations are based on the subjective feedback from MNOs and experience with implementation costs of similar systems, and present rough example estimation for a single MNO with variables presented in Table 7.

Table 7. Description of the MNO used in the scenarios.

Size of the MNO / amount of data chunk to process	10M domestic subscribers, 15B (10^9) monthly events generated for inbound, domestic and outbound data. 12 months of data is processed during the update (180B events). Estimated size of the data chunks: • 1 month – 630 GB (ASCII), 230 GB (binary); • 12 months – 7.4 TB (ASCII), 2.6 TB (binary).
Allocation of the processing resources	Two scenarios: maximum (full processing chain) and minimum (only extraction and preparation).
The number of processes to conduct by the MNO	All tourism forms, two scenarios: maximum (full processing chain) and minimum (only extraction and preparation).
Maximum allowed latency	Monthly updates. 15 days latency allowed (e.g. results for November, received by December 1st and processed by December 15th). For minimum extent: MNO has to extract, format and deliver the raw events data to NSI within the 15 days after the new month has begun. For maximum extent: MNO has to extract, format and conduct all data compilation processes including the aggregation and estimations within the 15 days after the new month has begun.

Out of these, the most cost-sensitive variable is the allowed / required latency time. Within this allowed period, MNOs have to be able to perform following processes:

1. Extraction of the data from internal registries;
2. Quality check for possible errors in the data (missing data, format, duplicates, location attributes etc.) and possible correction of the initial data;
3. Preparation of the data (including or excluding probabilistic geographic distribution) for further processing;
4. Relaying the prepared data for further processing system (end of the minimum extent scenario);
5. Frame formation and data compilation processes;
6. Aggregation and estimation.

During the latency periods all abovementioned processes have to be conducted and possible recalculations due to errors in the data have to be included. Table 8 describes the estimated costs of the two scenarios for a single MNO.

Table 8. Description of the maximum and minimum extent of implementation for an example MNO (in euros).

	Estimated cost maximum extent		Estimated cost minimum extent	
Pilot phase. Initial one-time data extraction to assess the methodology and usability (one time cost)				
Internal legal expertise	15 000	(15 days)	15 000	(15 days)
Data extraction and preparation for following processes	30 000	(60 days)	30 000	(60 days)
Hardware cost for one-time process (12 months of pilot data)	25 000		12 000	
Applying methodology on pilot data (manual process, no automation)	35 000	(70 days)		
Estimation	8 500	(17 days)		
Analysis of the results (coherence, quality issues etc.)	8 000	(16 days)		
Project management	13 500	(27 days)	6 000	(12 days)
Total	**135 000**		**63 000**	
Implementation of automatic data extraction and processing (one time cost)				
Automation of data extraction (monthly) and preparation for following processes (core system pushes the data to automated tourism processing environment)	20 000	(40 days)	15 000	(30 days)
Data storage hardware (data warehouse)	100 000		48 000	
Processing hardware	100 000			
Developing automated processes according to accepted methodology (frame formation, data compilation, aggregation, estimation)	250 000	(500 days)		
Exporting data automation	15 000	(30 days)	20 000	(40 days)
Quality control processes, including central system monitoring	20 000	(40 days)	15 000	(30 days)
Project management	46 000	(92 days)	7 500	(15 days)
Total	**551 000**		**105 500**	
Maintenance of the system (yearly cost)				
Quality manager (will correct the data if something is wrong)	40 000	(80 days)	25 000	(50 days)
System maintenance (SLA)	111 000		22 000	
Project management	7 000	(14 days)	4 000	(8 days)
Total	**158 000**		**51 000**	

Depending on the size of the MNO and the required latency, the costs of the implementation obviously rise (see Figure 13, Figure 14). The cost of shortening latency time might increase exponentially (latency requirement shrinks towards the near-real-time) as substantial resources are needed to process the data, in addition to the fact that maintenance requires the constant monitoring and attention of specialists.

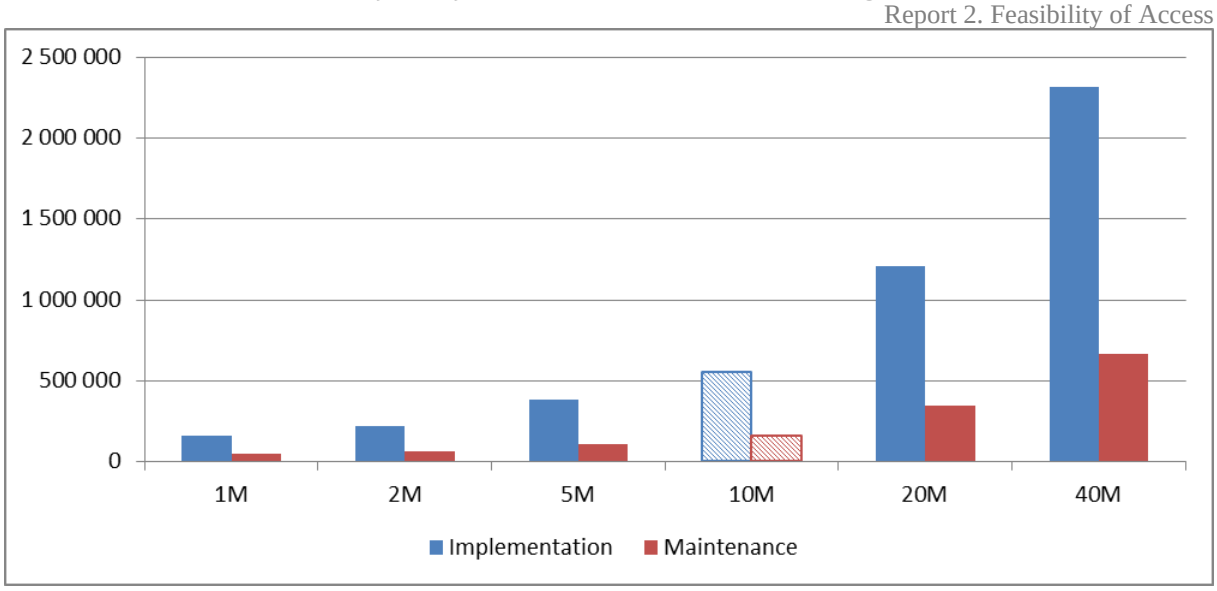

Figure 13. Cost of the implementation and maintenance of the system based on the size of the MNO (in euros, maximum extent, latency 15 days). Patterned bars represent the example MNO provided above.

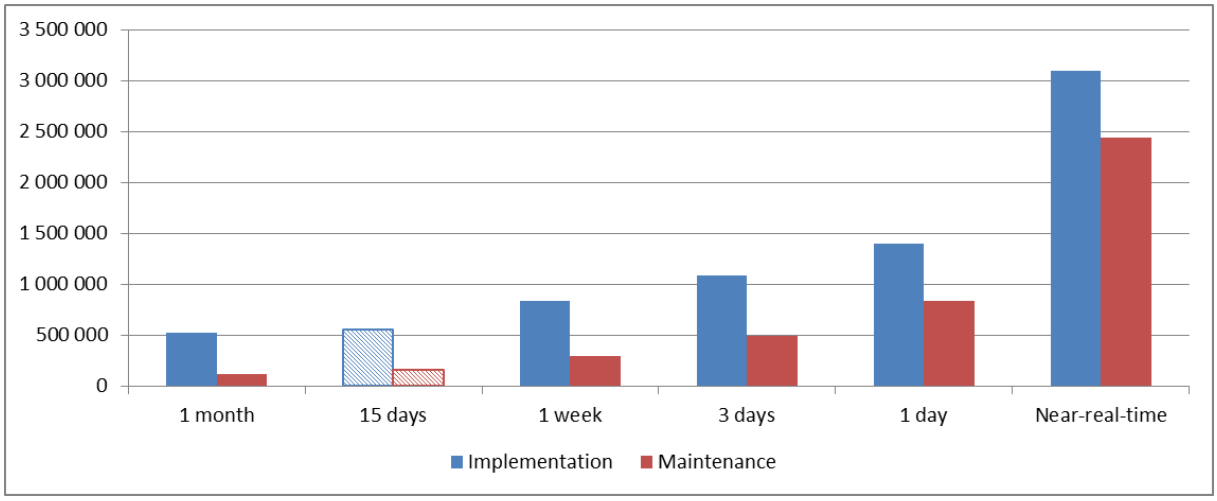

Figure 14. Cost of the implementation and maintenance of the system based on the latency of the processes (in euros, maximum extent, 10M size MNO). Patterned bars represent the example MNO provided above.

The burden on each MNO for providing data continuously might vary greatly depending on the size and internal system complexity of the MNO. It might require a number of different specialists and possibly additional hardware and software in order to implement a continuously working system. For shorter latency times, dedicated specialists have to be hired to monitor the system and take action for correcting any data or system errors. The cost might also depend on the complexity of procedures conducted within the MNO's internal system.

The total cost of the system comprises the implementation and maintenance costs on the MNOs side and additional costs for the receiving party (i.e. NSI). The total cost of the system (on the MNO side) is smaller with the minimum extent scenario as there are fewer

separate parts to the system to implement and maintain (see Figure 15). The cost of the system on the receiving side is presented in Report 4 Section 3.2.

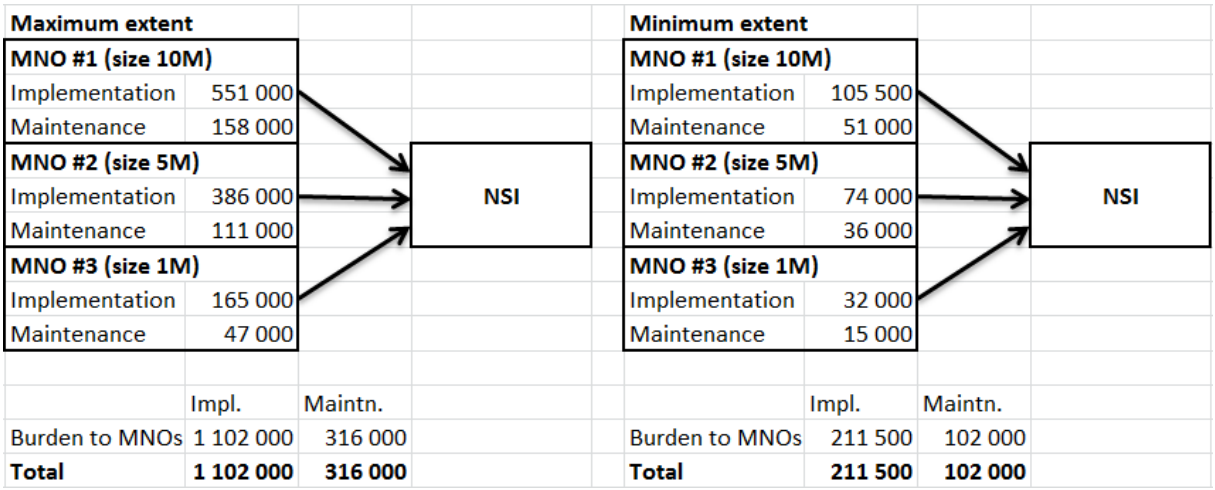

Figure 15. Estimation of costs (in euros) of two different scenarios with three local MNOs based on the example calculation. Sizes of MNOs are respectively: 10, 5 and 1 million subscribers.

The cost of the pilot project, implementation and the maintenance presented here is indicative and should not be taken as a realistic basis for expected burden. Each MNO will be able to provide more accurate figures that can be lower or higher than indicated here.

The possible cost dimension of statistics authority and the aspects of cost-efficiency will be discussed in Report 4 of the current study.

5.2. Business Secrets

Business secrets are strategically important information about the activity of MNOs. Access to such information by other MNOs could potentially harm the competitive advantage. MNOs consider losing sensitive business secrets to competitors more of a loss than any gain from learning about business secrets from other MNOs. MNOs are mostly concerned about the possibility of acquiring the following information by competitors:

- Number of subscribers (both domestic and roaming);
- Number of service activities (calls, messaging, data) in the network;
- Any information on the constitution of the subscribers (number of pre-paid vs. post-paid cards, socio-demographic information of the subscribers, number of subscribers from various foreign countries, etc.);
- Number and locations of network antennae (the release of this information might also be prohibited by law in some countries due to the terror threat to the vital telecommunications infrastructure of the country);

- Any financial information and strategic plans of MNOs;
- Technological capabilities and information on infrastructure and systems.

This issue is important as it was repeatedly mentioned by MNOs, but can be resolved by determining the procedures that do not allow such information to be exposed to competitors. Either such data would not leave the premises if the infrastructure of the MNO – only aggregated results provided to the receiving party, or the data receiver should have strict data processing rules that eliminate the possibility for third parties to distinguish data about different MNOs.

However MNOs consider NSIs to be reliable partners and expect that sensitive business information will not be made available for third parties.

5.3. Public Opinion

5.3.1. Overall Situation on the Public Opinion of Telecom-Owned Data

MNOs see the potential in the usage of their data. The tourism statistics domain is mostly comprehensible and the value of the data for the state is understandable. Yet, MNOs have pointed out the practical issues that might affect their decision to provide delicate information. In addition to legislation, the main problem they see is the potential decrease of their corporate public image and the decrease of trust of the subscribers, which might lead to a fall in the number of subscribers. It is a business risk and a barrier. Even if the data is handled following all privacy preservation techniques and rules, it does not mean that this is easily presentable to the public and customers. This chapter opens the discussion on the concerns of people with the use of their private data, how they might respond to threats to privacy and how to mitigate very emotive negative responses. It is important to deal with the media before it deals with you. A proactive, open and stakeholder-engaging approach can lead to positive results.

It is important to start with what concerns a person enough to voice their concerns or boycott a service. The 2011 Eurobarometer survey showed that 70% of EU citizens are worried about the misuse of their personal data (Special Eurobarometer 359). And the organisations and agencies they trust the least with their data are telecommunications and internet companies. From the Eurobarometer study as well as from the rest of the chapter below it can be inferred that people are afraid that the information is used for unknown or unintended purposes – that there is tracking undertaken with results that people are not aware

of, that they cannot control and do not benefit from. The citizens of Southern and Western Europe were, on average more concerned than those of Eastern and Northern Europe.

The fear is not directed towards location-based services (LBS) per se. In fact, more than 70% of US smartphone holders use some form of LBS (PEW, 2013). Over half of the world's consumers are willing to share additional personal information, such as their location, their top five Facebook friends' names and information about family members, in return for financial rewards or better service (Coleman Parkes, 2013). Again, Western European citizens are more concerned.

The literature around information privacy has commonly discerned a few types of concerns people have with regard to their sensitive information (see Table 9). First, the mere fact or possibility that data collection and retention takes place might concern a person. Second, they might be concerned about unauthorised parties gaining access to their data. Third, there is the fear that somebody might make an error in their data. And last, that the data is used in a way that was unknown or unauthorised by the person, e.g. by a third party for secondary use. (Smith et al, 1996)

Table 9 With regard to privacy concerns one can distinguish three archetypes of internet users, four types of concerns and six possible responses to threats to the privacy of personal information

Typology of Users by Approach to Privacy				**Concerns for Information Privacy**		**Information Privacy-Protective Responses**
The kinds of attitudes people have towards privacy issues				The concerns people have		How people respond to privacy threats
Traditionally		*Online*		Collection		Refusal to provide information
Fundamentalist	25%	3%		Improper access		Misrepresentation of information
Pragmatist	50%	81%		Possibility of errors		Removal
Unconcerned	25%	16%		Unauthorised secondary use		Negative word-of-mouth
						Complaining directly
						Complaining to third parties
Sheehan, 2002				Smith et al, 1996		Son & Kim, 2008

Publicly expressed fears associated with data retention and analysis are not unlike those that spring up in the debate around surveillance. Some, beginning with the sociologist Michel Foucault (1995), believe that in addition to its obvious function, surveillance also manages to create in everyone a feeling of always being watched. This so-called feeling of living in a panopticon, where you know you are being watched but unsure when or by whom, creates a sense of paranoia. The surveilled person, Foucault suggests, is always 'the object of information, never a subject in communication'. The expression of the 'I' is restricted, thus heightening the potential for inducing anticipatory conformity in the population (Maras,

2012). Indeed, the more individualistic countries (e.g. Western Europe, the US), where the 'I' that people take years to build is seen as more important, are in general more privacy-conscious (Milberg et al, 2000).

The question in the public eyes is, at what point is data analysis being applied inappropriately, or negatively for some individuals or groups? What is 'appropriate' in the access to somebody's personal information, does depend on the social context and the context-specific informational norms. If access to a person's information creates a ripple in their lives, then an infringement of contextual integrity is faced (Nissenbaum, 2010).

In the context of public opinion it should also be noted that fear is perceived, not always rational. The social production of fear does 'not take place through an individual's own experiences but through experiences of the others, circulated either in face-to-face conversations or in the media' (Koskela, 2009). Whether a threat to their person exists or not, people may resort to responses proper to a point-blank meeting with danger, and this 'derivative fear' becomes self-propelling (Bauman, 2013).

Whether rational or not, there are certainly people who would take action in case they feel their privacy is in question or their data might be used inappropriately (e.g. see the typology of possible protective responses in Table 9). To take a survey of app users in the United States, more than half of smartphone users uninstall or decide not to install a phone app based on concerns about the sharing or collecting of personal information (PEW, 2012). The wealthier, younger and more educated they are, the more likely they are to do so, i.e. be active in managing their mobile data. In a more frightening result, a poll in Germany showed that, as a result of data retention undertaken by MNOs in accordance with EU Directive 2006/24/EC (DRD), half of Germans would not contact marriage counsellors, psychotherapists or drug support services through telephone or e-mail (Forsa, 2008). Another poll of 2,176 randomly sampled Germans found in 2009 that 69.3% oppose data retention, making it the most strongly rejected surveillance scheme of all, including biometric passports, access to bank data, remote computer searches or passenger name record retention (Infas, 2010). Among those polled, 6.4% claimed they had switched operators to those that do not retain their call data (an additional 53.4% intended to do so). To an MNO in a country the size of Germany, considering it was a representative sample, public fears would amount to losing several million clients.

Not surprising then that a concern for its reputation drives MNO behaviour, including business decisions and investments. Reputation is a major driver to increase a telecom company's information security expenditure (PwC, 2013). The fears for company reputation

and of criminal theft were somewhat higher among the MNOs in North America than in Europe. Nevertheless, reputation management can also lead to cautious behaviour. The results of the survey show that there are many MNOs that strictly prohibited the association of their name with the data and results in order to curb the bad publicity. There are others that have allowed it but have taken a cautious step-by-step process in providing data, gradually testing the reaction of the public.

The NSA PRISM scandal has shown that uncovered secretive uses of private data, such as undercover surveillance programmes, come at a price not just for the state reputation, but also in tangible costs to private companies in that state in the form of cancelled contracts. Estimates for costs to US companies of such freeze-outs run as high as 35 billion USD (Rothkopf, 2013). Even worse, the events 'may turn back the tide of increased access to information that the information revolution was bringing' and bigger public adversity towards use of personal data, even if it is purportedly legal. In the aftermath, any new cases about the use of mobile positioning data might lead to a more serious reaction from the public.

News and media have been and still are the greatest influencer of public opinion (Page et al, 1987). Yet, in the recent years the definition of media has expanded, to the point that digitisation has certainly made control of information by institutions more difficult. In other words, for roughly a decade, governments have demanded transparency from their citizens, and so have corporations from their customers. People's data had to be available for whatever higher purpose deemed by the political and corporate powers. What document leaks like Wikileaks or the Snowden revelations reveal is the demand for a two-way street in that respect (Cardon, 2010). Transparency goes both ways.

That is not to say use of positioning data couldn't also influence public opinion positively. Mobile positioning data methodology helps better understand the impact that certain activities have and whom they might positively or negatively affect. A more informed decision-making could allow those decisions to be better explained and garner more relaxed opinions (Ahas et al, 2005). The question of appropriateness remains and the public seems to demand answers to it at any point they hear of personal-data-driven initiatives.

The survey undertaken for this project attests that organisations that use or might use mobile location data for tourism statistics are well aware that privacy is a major concern. MNOs have sometimes experienced many fears in the case of 'paranoiacs' and low consciousness in the case of 'optimists'. Comment from the survey: '*It can be scary to imagine a knowing-everything-government, especially about foreigners due to the political context. How could we ensure against abuses while tracking foreigner's mobility and not only*

tourists?' The press and the more PR-seeking individuals in the DPAs can react when data protection issues are even only slightly affected.

The privacy aspect should not only be managed methodologically, but also in the eyes of the public and the media – the mover of public opinion. Methodological correctness does not automatically lead to desirable publicity. The next few cases show the actions of companies and organisations, some that have drawn upon themselves negative publicity through the use of location data and some that have managed to successfully avoid such opinions, even garnering good press.

5.3.2. Negative Cases of Public Opinion

5.3.2.1. TomTom HD Traffic

TomTom witnessed 'unforeseen' use of its data by the police in the Netherlands. Data that the navigation systems company thought would be used to relieve traffic congestion and improve safety of the roads was also very useful for setting up optimal locations to catch speeding violators. After the discovery, in April 2011, and a public backlash the company apologized and had to redo the licensing contracts it had with the police (Williams, 2011b). The CEO made an online video statement saying they don't like what the police in The Netherlands are doing because their customers don't like it (TomTom's CEO…).

Yet, a month later, the company announced that it was planning to sell aggregated customer information to the Australian Roads and Traffic Authority, which could also potentially be used for targeted speed enforcement (Moses, 2011). This prompted the Australian Privacy Commissioner to say that companies that sell GPS devices should be very upfront about what they are going to do with information collected from the devices. Another expert on Cyberspace Law and Policy called on TomTom to authorise an 'independent technical analysis' of its data collection practices by an outside authority. These are suggestions that not only serve to protect the consumers but also the company itself from (unwarranted) criticism. At the end of one article, the Australian readers could vote whether selling aggregated data was an acceptable conduct of business from TomTom, to which 91% answered 'No' of the 8923 total polled (Moses, 2011).

Apart from the outrage received from privacy advocates, TomTom has a large online community and garnered a lot of fierce opposition in the forums. Some of the members of that community clearly stated selling data is the reason they switch to competitors. The company states the data is provided on an opt-in basis. Users found that opting out comes with a trade-

off – it removes the person from using the Live Traffic service themselves. Because the customers are forced to make a trade-off, they sometimes trade their service provider in for a competitor.

TomTom continues to use data from 80 million mobile phones and a million GPS devices in cars as sources for traffic information (Macura, 2011). The HD Traffic service, as it is called, is open in several countries, including The Netherlands, Germany and Austria. Many other players in the car navigation industry also have real-time traffic reporting capabilities, it having become a normal (auxiliary) use of telecom data by now. Perhaps the negative publicity could have been avoided had the licensing contracts clearly defined uses of the data from the start and had those terms been communicated externally, together with relevant analysis from external supervisory bodies. Acting in the dark and unforeseen uses can attract public outcry as well as lead to data owner's disapproval, possibly resulting in data use being blocked.

5.3.2.2. Apple Tracking 'Bug'

In 2011, it was discovered that Apple's iOS operating system (powering the iPhone, iPad and iPod Touch devices) collects data on the user's position and stores the information on a file on the device, then uploads it to Apple every 12 hours (Williams, 2011a). An application quickly appeared that visualised the privacy implications of the file on a map. It was never communicated that Apple collects this data, even though Apple was asked by two Congressmen of the US Senate to disclose location-data-collection practices (Chen, 2011). This lead to a lot of bad publicity in numerous newspaper articles and technology blogs; speculations, almost entirely negative, were abound for the week the company kept silent before posting a statement on their website (Richmond, 2011).

In the long-awaited response, Apple said they reduced the retention period on the phone from a year to a week and argued the data was necessary to reduce the time needed for location-dependent apps to function (Apple, 2011). Yet, the explanation did not thoroughly convince everyone that privacy was being uphold due to the nature of the file trace being left – the unencrypted file could easily be exploited through phone hijacking or loss of the mobile device. In addition, the opt-in process to share the data with Apple has been dubbed 'highly misleading' for the users (Hypponen, 2011). Google's Latitude application performs a similar task on Android phones, but unlike its competitor's service, the iPhone tracking file is not dependent upon signing a specific EULA or even the user's knowledge, but it is stated at the end of the 15,200 word-long terms and conditions of the iTunes program used to synchronise

the phone and the computer that 'Apple and [their] partners and licensees may collect, use, and share precise location data, including the real-time geographic location of [the user's] Apple computer or device.'

Many questioned the motives of the company and assume that the data is used for geomarketing and for its Apple Maps as well. Apple has a reputation of being secretive, yet user-focused in development, which might have helped mitigate the effects of this particular faux pas but it is doubtful it won them any new customers. The company admitted in its statement that 'users are confused, partly because the creators of this new technology (including Apple) have not provided enough education about these issues to date' (Apple, 2011). It is that confusion, and silence on the part of Apple to reveal its true intentions, that created good ground for speculation and fear-mongering in the media. It could have been avoided through 'enough education' as the company puts it. After all, in a subsequent iPhone release, Apple proudly pronounced the improved continuous motion and location-tracking abilities of a new processor powering its devices (Watson, 2011).

5.3.2.3. A1 Traffic Data Stream

Austrian mobile network operator A1 already brought location-relevant data services to their users in 2003 (mobilkom austria, 2003). In 2009, the MNO started a data stream service that provides aggregated location data of their subscribers to partner companies. The data offered valuable analysis material for geomarketing, market research and planning projects. The initial press release stated that 'Companies wanting to use the data for further analysis or their own applications may purchase […] unlimited access to these anonymised data' (A1 Traffic…). Data from A1 Traffic Data Stream was also processed and visualised by a third company. Users could find the latest traffic news or the current average speed on the Austrian roads real-time online at a web address that is now defunct. The MNO has claimed the service was stopped until further notice allegedly due to public opinion and data protection issues.

Arguably, the way the service was presented created too many questions in the public eye. For one, uses for the data and choice of partners were not limited, which suggests an openness to experiment. It was easy then to speculate on various abuses of the data, more so that access to anonymised (raw) data was mentioned. Secondly, creating very effective visualisations of mobile location data might have that affect, as was seen with the public opinion outcry in Germany, when a newspaper published the visualisation of a single person's movements based on just the mobile positioning data gathered on that person by the MNO

(Biermann, 2011). A1 followed their experience from a previously conducted art experiment, where a real-time map of the mobility of humans in the city of Graz, Austria was displayed for a limited time at an art exhibition (Ratti et al, 2007). The art project was presented well with good publicity. Still, the experience from art to business was not transferrable.

A1 has since returned to a closed model and uses the A1 Real-Time Traffic capabilities to power its A1 Navi navigation system's traffic warnings.

5.3.3. Positive Cases of Public Opinion

5.3.3.1. Regional Commuting Study in Estonia

The Mobility Lab of the University of Tartu composed a study about commuting in Estonia. Using call detail records for the years 2007-2009, home-work and home-leisure connections were mapped. The results of the study were used for a new regional development strategy and for mapping the major catchment areas of Estonia. The project was a success from both the outcomes and the publicity perspective, with continuous updates in the future and a planned development of a monitoring system (that promises automated updates of the results). Through being open from the start the researchers gradually created trust with both the public and the MNO, leading to a more stable cooperation.

This was not the first published project from the research team. In fact, the team has worked together with the MNO for some 10 years. The initiative came from the researchers' side. At every step in the cooperation, wherever another project was started, the MNO has asked to release public statements regarding the research undertaken. When the reaction from the public was positive, better and better access to data was granted. The process allowed the research team to build up trust towards the MNO and the society in a step-by-step manner. Whereas the start has been slow, in the end the gradual build-up of trust has allowed for easy access to data and even the possibility of automated updates.

The press campaign for the follow-up study was heavily covered in the Estonian media in 2013. The statements were clear about the benefits for the client - the Regional Ministry of Estonia - and the results (Government study…). The MNO partner was always clearly named in the Estonian-language press releases. The client was insistent on adding a visualised map of the results, which helped good publicity (interestingly, where for the A1 case, visualisation led to a negative public reaction). It was clear from the visual that results of this detail could not have been possible without the use of mobile data. The level of

aggregation was sufficient enough not to raise too many questions, although there were also the mandatory few negative public comments.

One should note that at the same time a newspaper published a subsequent article (Rajalo, 2013) about the many alternative possibilities to use mobile positioning data in research and business. There, the interpretations of the few commenters ran mostly to the negative. It does seem that imagination is an enemy to positive public sentiment when it comes to people's own mobile location data.

5.3.3.2. Google Maps' Crowdsourced Real-Time Traffic

Google announced on their blog in 2009 that they collect traffic information from anyone using Google Maps with GPS activated and use it to calculate and display real-time traffic situations on the roads of the US (Google, 2009). This supported other sources of information, as in official data from the road authorities. To respond to privacy concerns, Google stated they provide an opt-out option, only use data where it can be aggregated, and claim everybody's journeys' start and end points are snipped. The public got an 'aha' moment – an explanation to the source of Google's information about traffic- and the whole revelation can be seen as a successful public announcement. The one blog post pointed out all the benefits for the user (first), all the risks and how they are being dealt with, and even spun the story as a positive competitive edge over their competitor Apple, who at that point hadn't allowed to crowd-source traffic on the iPhone app.

The media reported the news widely and helpfully – as a service that employs the information of people to help all the users and at the same time has privacy issues 'under control' (Sorrel, 2009; Tofel, 2009; Cheng, 2009). Most articles ended with a hopeful expectation that the service to get increasingly better as more people join to share their data.

Google shared the news about the crowd being one of their sources of information a few weeks after the Live Traffic service had come live. The timing of it was positive, as the benefit of the service had already had time to sink in. Google has since expanded its data acquisition by acquiring additional data from two of the biggest crowdsourced traffic information providers (Inrix and Waze), allowing it to scale the Live Traffic service to many new locations.

5.3.3.3. Better Aid after the Haiti Earthquake

In the aftermath of the earthquake in Haiti in 2010, many people fled the capital. Researchers from Sweden's Karolinska Institute and Columbia University in the US asked the

country's biggest cellular network, Digicel, to provide anonymised information about the phone towers that people were using. From that data they estimated that 600,000 Haitians had left the capital in the first 19 days, and were able to locate concentrations of displaced people on a map. This information was already useful for aid workers, who could target their supplies to specific locations. But when a cholera epidemic broke out, the researchers were quick to react. In just 12 hours, they could now assess, which areas had received refugees from the cholera outbreak zone. (Bengtsson, 2011)

The paper released by the researchers received positive press (Mobile phones…; McNeil, 2011) and demonstrates that analysis that has already been undertaken on the movement and migratory patterns of populace can prove extremely handy in the event of a crisis, such as disease outbreaks. The fact that the work was peer-reviewed added to the credibility of the researchers' best intentions and to the soundness of the methodology, also from a privacy perspective.

5.3.3.4. Telefonica Smart Steps

Telefonica created a whole new business unit to deal with big data. The company's Dynamic Insights unit's Smart Steps programme provides anonymised aggregated data and they make it very clear through pilot studies and case studies how the data benefits its clients. In the initial press releases Telefonica painted a very good privacy picture by explaining how the data in question is solely in aggregate form (Telefonica, 2012). An added boost is the approval of the UK Information Commissioner's Office through a report on BBC – the data cannot be linked back to any individual person (Telefonica hopes…).

In addition, another press campaign from the Dynamic Insights business unit focused on the work of the data scientists at Telefonica for the greater social good. The example given was an avian flu tracking project undertaken that has implications for disease tracking worldwide (Oliver, 2013). Among others, they also collaborate with the Open Data Institute to organise events where participants are asked to programme data-driven socially beneficial applications over the weekend (Telefonica, 2013). Again, a display how the company engages well-respected third parties and is confident about the non-traceability of its data.

For Telefonica, public opinion was managed through a coordinated plan: step-by-step adoption starting with coarse aggregate data, approval of independent data protection officials and bringing out the social good of the data science. All in all, the openness and engagement with an embracing attitude towards external parties leads to the company being well accepted as deriving a net benefit from its data. The show has paid off.

5.3.3.5. MIT's Senseable City of Rome

The MIT Senseable Cities Lab depicted the urban dynamics of the inhabitants of the city of Rome through visualisations of mobile phone data, demonstrated at the Venice Biennale art festival of 2006. It followed a successful smaller art experiment in Graz, one of the very first projects of this kind (Ratti et al, 2007). The project in Rome was a show of the possibilities of mobile data on urban planning and transport when location data points of hundreds of thousands of people are aggregated and analysed. Visualisations at the exhibition and in the press were very impressive, even for that time.

The press was overwhelming and positive, from the local papers to the Financial Times (Balbi, 2006; Waters, 2006). Every press release and article also considered the privacy issues that might arise, something the research team always acknowledged and overthrew. To protect privacy, the researchers started out with aggregated data, already grouped and stripped of personal identifiers. But the researchers agreed that people should always retain a certain amount of control over how their data are used. The project instigated a follow-up to improve such mapping of Rome (MIT, 2007). In addition, the local authorities and the transport company of Rome later used it as an example of how to improve their OD matrices and overall business. In fact, the expectations were driven so high that people are anticipating outcomes from the real-time city projects and feel disappointed when their expectations are not met (MacManus, 2009).

The MIT's Senseable Cities projects in both Graz and Rome were good examples of an effective way to demonstrate the possibilities of mobile positioning data to the public. They created a swathe of positive opinions and very few doubtful statements, most probably because of the researchers' thorough work with the press. Although concerns must have been raised, they were not accentuated by the media. If anything, the show left people expectant of the benefits.

5.3.4. Conclusion and Recommendations on Public Opinion

It has been established that fears of tracking might be rational or irrational, but they are driven by the fear of the unknown – a sense of being watched but without knowing by whom or for what purpose. Media can be very fast to react and negative press on the (mis-)use of sensitive data sources creates more adversity among the populace, driving some people to stand up in anger or disappointment regardless of the purpose and if the data was used appropriately. The underlying question is whether data is actually used in a manner that is appropriate and the reasons understandable and acceptable to stakeholders involved. This

prompt of appropriateness should be tackled with care early on, in a way that it can be easily deduced that there is support for the measure among objective stakeholders, and there is little chance of negative surprises. The projects that opened up early on and created a spectacle, like the ones from MIT, Tartu Mobility Lab or Telefonica, were well accepted overall.

What can be learned from this? There are two main options to collect data for any purpose: a) enforced data collection by the state (using a statistical act or other instruments); b) voluntary provision of the data for research or commercial services (that can also be used by the state in the end). With option A, the public opinion vector is mostly towards the state and MNOs are targeted indirectly as forced data providers. With option B, it is more complicated and may require more work. The example of TomTom shows that negative press from unforeseen uses can put a relationship with a data donor at risk, even if the data was used to foster public safety. What is more, the case of Austria Telekom (A1) in Austria shows that although an MNO uses anonymous data from its subscribers in commercial service according to the law and all principles of privacy protection, the reaction (or the fear of) from the public may force the MNO to discontinue the service. The responsibility might be shifted towards the state but public concern can affect the continuity of data streams by the MNO if the sentiment is not managed properly.

With both options A and B it is suggested to conduct several activities in order to ‘play it safe’ regarding public opinion:

- Preventive public announcements, campaigns and a simple explanation for the public of why and how the data is used, and what are the benefits to the public.
- External communication should define what the data is definitely not used for and leave the public and the media with as little freedom as possible to come up with their own interpretations. Establishing boundaries will hopefully remove some derivative fear leading to fewer opportunistic (new) media articles and a ‘softer’ public reaction;
- MNOs, data processors and data consumers strictly following the legislation and privacy protection techniques with the involvement of a national data protection agency, data protection NGOs, media, etc., and have an open discussion between the interest groups;
- Statements from external stakeholders, especially those involved in data protection issues, should be included in announcements and campaigns;
- The licensing contracts should set boundaries for the uses of data, so as to limit unforeseen uses that could scandalise the whole campaign;

- Involving as many MNOs as possible in the data acquisition process to eliminate the risk to a single MNO;
- An option for the MNOs to leave their involvement anonymous.

Public opinion is one of the key issues concerning the usage of the data. Even if the legislation and methodology are in place, regulations are followed and the overall purpose does not aim for tracking individual people, public opinion may force decision-makers (politicians, MNOs) to withdraw as the negative effect in society might exceed the practical value of the data. The cases presented here and the suggestions help prepare users of mobile positioning data to deal with the public.

5.4. Benefits from Providing the Data

There are two mentioned options for providing the data – a) enforced data collection by the state (using statistical acts or other instruments); b) voluntarily providing the data for research or commercial services (that can be used by the state). With the first option MNOs are required to implement the system for data extraction and have to finance the implementation themselves (in most cases). Whether their expenses are reimbursed by the state or not depends on the situation in the specific country and the budget of the statistical office. Neither option offers direct benefits to the MNOs. The MNOs suggest in the interviews that with the second option 'in the long run there would have to be a working business model for a possible cooperation concerning tourism statistics.' If the data is provided based on commercial agreements, MNOs can earn direct extra revenues.

MNOs can also benefit indirectly from the results of the tourism statistics as they usually include official statistics in their regular market share comparison. For example the MNO can assess its inbound roaming service market share for each country by comparing the number of inbound subscribers to the number of total estimated tourists on the country or regional levels. Currently this assessment is usually done based on the accommodation statistics which is insufficient. Same logic applies to the outbound roaming service market share where existing data for outbound travellers is even less accurate. MNOs have expressed interest in obtaining more accurate and granular data concerning the total inbound and outbound tourism that would have different classifications (duration of stay, regional level).

MNOs can also take advantage of the resulting data processes that are usually new and not known to MNOs. These results can be used internally or in some cases externally for commercial purposes. If the MNOs are required to process the data for the NSIs, the resulting

aggregated data can be considered as a by-product and there should be no limitations of how these results can be used as they no more hold the personal information on their subscribers. Although this is again a grey area, potentially such aggregated results can be monetized for other interested parties.

Internally, MNOs can use the resulting data in various geographical analyses concerning the location and movement of their subscribers. Alternatively to their customer relationship management information on the registered addresses of the subscribers, the aggregated map of the usual residences can show the actual home areas of their client base.

With the revolution of 4 generation of antennae (LTE) and revolution of services based on the location of the subscribers, MNOs can get viable information that can be used for planning marketing activities, development of network infrastructure (to places where there are more subscribers but weaker service level), etc. Such information can be retrieved from the aggregated data that is generated for tourism purposes (e.g. aggregated number of usual residence's, movement corridors of commuters).

It is not possible to provide specific list of benefits for MNOs as different MNOs value different aspects. Some MNOs only consider financial benefits and do not see any value from any kind of data processing, some are keen to participate in projects where society can benefit from their participation regardless if such participation carries any hidden agenda (commercial interests, supporting main business by helping government). There are different possibilities for MNOs to use the data internally, but as seen from the responses from MNOs, many of them value different benefits that are often contradicting.

6. Practical Experience on Accessing the Pilot Data

Survey results show that 86% of the respondents are aware of the possibilities of using mobile positioning data, mere 14% of all respondents actually use it. Among that minority, some use phone data using special applications installed on devices. Responses from the survey and interviews indicate the importance of trust and clear understanding in cases where data has been obtained and also personal connections put into effect, which makes the communication and interaction between different parties and organisations essential for a sustainable operation. The procedures needed to go through to obtain the data have similarities in the responses: long negotiations with the MNOs, agreements, non-disclosure agreements; the licenses for the data; special software implementations for processing the

data. Usually the data from single MNOs is used and often the sample is used. In some cases respondents were not allowed to name the MNO that provided the data.

Interviews with MNOs revealed that they are mostly interested in sharing data. If at all, they currently provide data mainly to public authorities based on official requests or to data brokerage partners. The MNOs interviewed (in France, Germany, Finland, Belgium, Netherlands, Estonia and UK) had dedicated teams or partners to handle the data business. Due the international operation of most MNOs and the international structure of the 'data teams', the interviews did not only reveal the national activities and strategies of MNOs concerning their data business in the abovementioned countries but also cover most other European countries and even overseas activities. The interviews with the Telefonica Dynamic Insights people in Germany and UK, for example, gave insights in Telefonica's global activities in the data business (http://dynamicinsights.telefonica.com/479/about-us). The same goes for the interviews with the other internationally active MNOs, such as Orange, Vodafone, Deutsche Telekom etc.

Although the MNO does not provide individual localisation data to every asker in fear of legal ramifications from the personal data protection act, aggregated datasets can usually be offered after the approval of a competent national agency. It is a welcome revenue stream, but the impact on the company balance sheet must be positive. Even if the companies were willing to share pilot data below cost, there had to be a clear understanding about the future developments of the cooperation with Eurostat or NSIs in order for them to make the decision to invest cost.

During the current project partners tried to contact all MNOs from their respective countries and outside. Although the efforts to obtain the pilot data for the current study were great, only existing pilot data in Estonia was used. Other MNOs expressed their interest in participation, but the process of the access to the data was not completed by the time of the current report. The main reasons for that is an overall slow process and legal limitations. Several organisations with existing data were also contacted, but either they were not able to provide the data because of limitations of the agreement or the data was not usable for the methodology following the data scenario described in Report 3a.

6.1. Feasibility of Data Access in Estonia

6.1.1. MNOs in Estonia

At present there are three MNOs active in the Estonian market: EMT (TeliaSonera), Elisa and Tele2. EMT holds the leading position (46%) representing most of the social and geographical groups of subscribers in Estonia. Elisa and Tele2 share the rest of the market and tend to over- and under-represent different subscriber's groups.

6.1.2. Process/Efforts Undertaken to Access Data from the MNOs

EMT and Elisa have been partners of Positium and University of Tartu (members of the consortium conducting the current study) for several years. Analyses based on the mobile data in tourism, transportation and other public domains have been conducted by University of Tartu, Positium and some other organisations already for several years. Data from two MNOs are used in the current study through special agreement with MNOs. Data from Tele2 has been tested, but licence for the use in the current study was not obtained from the MNO by the time of the report.

6.1.3. Barriers of Access

6.1.3.1. Public Opinion/Legal

Trust and transparency have been the main reasons for the MNOs to be able to utilise the data. The projects that have gain value from the mobile data, have been publicly advertised and discussed with Data Protection Agency as well as publicised in the media. Although there has been negative feedback and limitations on some attributes of the data and as the overall topic of using the mobile data promotes the superficial comments, the credible feedback has usually been neutral or positive as the reason for using the data and the value gained are mostly clear and give no ground for speculations for intrusion of the privacy. See Report 1 Estonian use cases for further information.

6.1.3.2. Business

MNOs are looking into the possibilities of utilising the data they possess both for internal analyses as well as possible commercial and non-commercial usage. Several public projects have been conducted in academic sphere and in public interests. The commercial use of the data is still under consideration and is being investigated.

6.1.3.3. Technological

In cooperation with Positium, the data mediation system has been developed for extraction and processing of the mobile data into tourism statistics and other outcomes. The system extracts the data from different databases and registries of MNOs, tokenises it and processes with specific set of algorithms until the aggregated results and estimations for specific domains are produced. The results are suitable for use in monitoring systems, data exchange APIs, interactive applications, analyses etc.

6.1.4. Assessment of Success/Conclusions

Estonia has been a leading country in the use of the mobile positioning data in various domains. This is because Estonia is one of the leading IT countries in the World with e-government and various public and private IT applications that create the environment that support new technological development in many areas. University of Tartu and Positium have been working with Estonian MNOs for several years, developing the methodology and technological processes, with consultations with Estonian DPA.

6.2. Feasibility of Data Access in Finland

6.2.1. MNOs in Finland

There are three primary MNOs in Finland, whose combined market share is 98%. These MNOs are Elisa (39% market share), TeliaSonera (34%) and DNA (25%).

6.2.2. Process/Efforts Undertaken to Access Data from the MNOs

The objective was to contact all Finnish MNOs and discuss the steps necessary to acquire pilot data that can be used during the course of the project. The following steps were taken to reach this objective.

1. Identify the right contacts within the MNOs (Apr-May 2013). Statistics Finland had no existing contacts to the MNOs so the contacts were acquired mainly by A) contacting the MNOs directly and asking for relevant contacts; B) using contacts that were provided by Positium; C) consulting other agencies in the Finnish governmental sector that have run earlier projects involving Finnish MNOs.

In the case of all three operators, the right contact persons were found for each MNO. Typically these persons were located within the roaming department of each MNO.

2. Meetings with MNOs (May-June 2013). Meetings with all MNOs took place in May-June. At the meeting STATFIN and Positium presented the project and objectives and the opportunity to access the pilot data was discussed. The feedback from all MNOs was very similar. While they showed great interest towards current project, all MNOs stressed the need to have a statement first from the Finnish Data Protection Agency before they could proceed.

3. Statement request from DPA (19th June 2013). STATFIN submitted statement request to Finnish DPA on 19th June. In this request, the Data Protection Agency was asked to state whether pseudonymised raw (CDR) data constitutes as anonymous data and whether MNOs could deliver such data for STATFIN.

4. Statement from Finnish DPA (4th Oct 2013, see Annex 10. The Initial Responses of the DPA in Finland). The statement was provided by the DPA as late as October 4th despite several requests for speedier processing. The primary reason for the delayed statement was the holiday season during which none of the necessary persons from DPA's office were available. The English translation of this statement is available in Annex 10.

The main conclusion of the statement is that the described pseudonymisation does not guarantee anonymity of the data if the identifier individualising the subscriber connection is the same all the time, plenty of observations are collected from the subscriber connection by means of location data and timestamps, and the time of data collection and the number of observations accumulated in this way are in no manner restricted.

5. Interview with Finnish Data Protection Ombudsman (Reijo Aarnio, Oct 2013, see Annex 10). The head of Finnish Data Protection Agency, Reijo Aarnio was interviewed in person to get more in-depth knowledge on the statement and what could be possible in Finland based on current legislation. A translation of this interview is available in English in Annex 10.

6. Interview with Finnish Telecom Regulatory Authority FICORA (Nov 2013). FICORA is the main regulatory authority concerning the processing of telecommunications data. The main question presented to them was whether the MNOs are allowed to process telecommunications data for use in statistics.

7. Follow-up meeting with Elisa (Nov 2013). The meeting was held with Elisa again to discuss the outcome of the statement request and the way to proceed from there. Elisa clearly stated that a positive statement from the DPA is required. Furthermore, a business case should be constructed since no legal obligation exists for MNOs to provide the data.

8. Next steps. The planned next steps include a revised statement request to the DPA based on aggregate tables (pre-processed) rather than pseudonymous raw data. A positive statement would then be followed by business case negotiations with the MNOs.

6.2.3. Barriers of Access

6.2.3.1. Public Opinion/Legal

The current interpretation of the Finnish Data Protection Act is that the MNOs are not allowed to deliver raw data even if the data is pseudonymised. Furthermore, the current Statistics Act also does not constitute a legal basis for obliging the MNOs to provide such data. This is the main barrier of access.

The possibility to acquire aggregated (pre-processed) data from MNOs will be examined further. In this scenario, significant resources need to be invested by the MNOs in order to produce the necessary tables and for this purpose a feasible business case needs to be established with the MNOs.

The public opinion is a major topic in Finland considering the monitoring of people. The Finnish DPA stressed that a major challenge in this kind of projects is to communicate the need for such monitoring to the public.

6.2.3.2. Business

No estimate was discussed at this point in terms of cost for providing the data. The MNOs clearly see the value of taking advantage of their data for tourism statistics and other areas. Prior to engaging in a pilot study, a feasible business case needs to be set up. The feedback of one MNO also was that they would participate if the other two MNOs participate as well.

Concerning business secrets, one MNO commented that they consider the locations of antennae within their network a secret and would not disclose this information.

6.2.3.3. Technological

There should be no major technological barrier. The operators run similar systems as in Estonia where data has already been extracted successfully.

In case operators would provide aggregated tables instead of raw data, there is a question of expertise and resources required by the MNOs to run the ‘tourism algorithms’.

6.2.4. Assessment of Success/Conclusions

During the course of the work, Statistics Finland has established contacts to MNOs, the Data Protection Agency and the Finnish Telecommunications Regulatory Authority.

Considering the current legislation and interpretation of the Data Protection Act, the MNOs are not allowed to deliver the raw data even if the data is pseudonymised. The current Statistics Act would need to be updated in such a way that Statistics Finland would have the legal right to obtain this data from the MNOs. This is the main barrier of access presently in Finland.

Based on the work done so far it's too early to make the final conclusion on whether the MNOs could provide the data in the form of aggregated tables. To make this final assessment, another statement request will still need to be made to the Finnish DPA followed by negotiations with the MNOs on the terms of a voluntary data provision. This work is due to be continued by Statistics Finland even if the matter will most likely be fully resolved only after the project is finished.

6.3. Feasibility of Data Access in France

6.3.1. MNOs in France

Mobile telephony is structured around two main types of mobile network operators: conventional operators (with their own mobile network) also called MNO (Mobile Network Operators) and MVNOs MVNO (Mobile Virtual Network Operators), e.g. in France:

- four standard operators: Orange, SFR, Bouygues Telecom and Free Mobile;
- forty mobile virtual network operators, known as MVNO, the networks that use the previous.

The market share for these four MNO (is 91.0%) in 2013: Orange (36.6%), SFR (29.0%), Bouygues Telecom (15.2%) and Free Mobile (10.2%).

6.3.2. Process/Efforts Undertaken to Access Data from the MNOs

The objective was to contact French MNOs and discuss the steps necessary to acquire pilot data that can be used during the course of the project. IFSTTAR had no existing contacts to the MNOs so the contacts were acquired mainly by:

a) contacting the MNOs directly and asking for relevant contacts;

b) contacting people that have worked with such data and ask them their contacts names / mail / phone for the MNO.

It took some time to reach the right persons at two main MNOs (Orange, SFR). During the exchanges to conduct the interviews, Orange stated that they cannot provide raw nor pseudonymised raw data. Only aggregated data may be provided. At current stage there are still negotiations going if the pre-processed aggregated or only aggregated raw data could be accessed.

6.3.3. Barriers of Access

6.3.3.1. Public Opinion/Legal

The presence of a given fine location provides the possibility (using external additional registry) to identify the subscriber using pattern matching algorithms. Such opportunity can potentially be used by MNOs or data processors. However National Institute of Statistics deal with personal data on everyday bases, so this is not a precedent. National Institute of Statistics is in the exemption list who are allowed to conduct the indirect collection of the personal data. The National Institute of Statistics must obtain the consent of all concerned parties (except the subscribers) in order to start collecting such data.

Another concern is the ownership status of the data. On one hand the stored data for specific subscribers is their property and the licence to use is given to MNOs. The formed databases that hold the data are owned by MNOs.

There are two possibilities for NSIs to get access to the data: agreements with each MNO to provide the data or adopt a legislation that requires operators to provide data to the NSI. This may be a national law or a European regulation.

6.3.3.2. Business

French MNOs have been exploring the possibilities to utilise the data. Even in tourism statistics some aggregated form of data is being offered to municipalities (14 000 €). The MNOs clearly see the value of taking advantage of their data for tourism statistics and other areas.

6.3.3.3. Technological

There should be no major technological barrier. The operators run similar systems as in Estonia where data has already been extracted successfully (The French Riviera Tourism

Board have tested the data from Orange to analyse the reliability of such data). In the case of having aggregate data by different operators, the risk is to compile different and non-comparable aggregate data, if the post-processing algorithms are not the same.

6.3.4. Assessment of Success/Conclusions

During the course of the work, IFSTTAR has established contacts to MNOs, the Data Protection Agency and the National Institute of Statistics.

The Data Protection Act requires the data controller to conduct preliminary formalities to the National Commission on Informatics and Liberties (CNIL – the French Data Protection Authorities), in this case, a normal statement is sufficient and collect the consent of the persons concerned. However, if the data is anonymised at short notice, consent may be replaced by streamlined information of people, limited to the identity of the controller and purpose of the processing. It is clear that if location data is collected directly in an anonymous form, the data cannot be classified as personal data and the Data Protection Act does not then apply.

6.4. Feasibility of Data Access in Germany

6.4.1. MNOs in Germany

At present there are four MNOs active in the German market. The distribution of market shares in 2013 is: Deutsche Telekom 32%; Vodafone 30%; E-Plus 21%; O2/Telefonica 17%.

On July 23, 2013, O2/Telefonica and E-Plus publicly announced their plans for a merger of the two providers in Germany. The combined customer base of the two providers would reach 43 million subscribers and constitute rank 1 in the German market (please also see Report 1).

6.4.2. Process/Efforts Undertaken to Access Data from the MNOs

MNOs, in Germany as in other countries, are huge multi-department, multi-national enterprises with ten thousands employees and several business locations all over the country. From the outside, it is not easily detectable if, how and who is handling their mobile positioning data. This makes it a complicated and time-consuming practice to get in touch

with the right persons within the MNOs to talk about accessing their mobile positioning data. The following describes the steps that are necessary to finally acquire this data.

1. Identification: In order to identify the right contact persons in the four MNOs, parallel paths were taken: a) calling their headquarter/send e-mail to the central address (e.g. info@...); b) conduct web search for the persons/departments handling/marketing mobile positioning data; c) using professional and personal networks to identify the right persons. In the case of Germany, the third path lead to success (i.e. ask someone who knows someone within a MNO to provide the right department and contact details). It took around 6 weeks to get the contact details within all four MNOs.

2. Get in contact: Right after receiving the contact details, the four MNOs were approached, first by e-mail. All MNOs have departments/start-ups that try to exploit new business opportunities by selling insights based on the different kind of data the MNOs could possibly provide. These were the departments that were dealt with. The feedback differentiated between the four MNOs: One answered instantly and a detailed discussion by e-mail followed before agreeing on a meeting; one answered after a while, one after two e-mail reminders and both asked for a meeting; the last (E-Plus) was not interested in the study (as they do not market their mobile positioning data and as a merger with O2/Telefonica is planned). It took 4 to 12 weeks from the first e-mail until the date of the meeting.

3. Meeting: All meetings took place in person at the premises of the MNOs with two to three representatives of the MNOs. The results of the discussions are covered in Section 6.4.3. The duration of the meetings was between 1.5 and 3.5 hours.

4. Stay in touch: The MNOs were kept updated about the project progress and urged the need for the pilot data.

5. Acquire pilot data: Because of Eurostat's high expectation towards the data quality (comparable to Estonia), the complex data protection situation, the cost involved for technical processing of it was not possible to acquire any German pilot data for the study (also see Section 6.4.4).

6.4.3. Barriers of Access

6.4.3.1. Public Opinion/Legal

Data protection: Is nowadays the biggest barrier of access in Germany. The legal basis is set by the EU/German legislation (see legal expertise in Section 3). In the interview, the regulation authority (Bundesbeauftragter für den Datenschutz und die Informationsfreiheit

Referat VII Telekommunikations-, Telemedien- und Postdienste) put forward three general rules, based on the German legislation: (a) Everything that is not explicitly allowed is forbidden; which means: regulation authorities approve/disapprove projects in case by case decisions. (b) Appropriation: MNOs are only allowed to use 'personal data' to conduct the services that are agreed on in the contract with the subscriber (i.e. provide telecommunication services and billing for these services). (c) In order to use mobile positioning data outside the appropriation, the data needs to be anonymised. Pseudonymisation is not sufficient. 'Anonymisation' by the definition of the German regulation authority is reached, if the effort to trace back the anonymised data to the original person is immoderate and economically not viable.

At present, the 'anonymisation key' for mobile phone data has to be changed every 60 minutes (e.g. in traffic applications). All MNOs are working on new, complex anonymisation technologies which they would like to have patented, certificated and finally released by the regulator.

Public opinion: Is a big issue in Germany. No MNO would risk its good reputation because of data projects. But all MNOs see the study as a possibility to get good press (with Eurostat and a rather scientific consortium behind it). It would especially be necessary to show that thanks to big data, additional value can be generated for the public. There have recently been two quite well-balanced reports in big German magazines about the potential benefits and dangers of big data ('Leben nach Zahlen' 2013; 'Big Data...' 2013), on the other hand the NSA/Snowden affair has probably not positively affected the public acceptance of using anonymous mobile positioning data.

6.4.3.2. Business

Financial: MNOs are private business oriented enterprises in a heavily competitive market. In the end, they are only able to prioritise subjects when they have a good effect on the balance sheet. They would invest their own money only if they see the business case. Even if they are ready to supply pilot data for free, they would need compensation for the work related to providing data (quote: '30-50.000 EUR' to process pilot data). However, the Eurostat study is seen as a possibility to set a good example with the help of the public impact. This could smooth the way for later and stronger commercial projects.

Business secret: No MNO is willing to provide raw data. All know-how/processing of the aggregation process is to stay within the MNO.

6.4.3.3. Technological

Generally, there seems to be no technological barrier. But of course, technological issues do have an impact on implementation and running cost. To identify technological needs and the respective costs, the MNOs need to have precise definitions of what kind of data is searched for. The cost part is especially eminent, as MNOs are under no circumstances willing to provide raw data – only aggregated data in form of reports. In the end it is a question about the distribution of the full processing chain.

6.4.4. Assessment of Success/Conclusions

The German MNOs are very interested to market the use of their mobile positioning data and are thinking about products. Presently, there are no products available yet.

The main barrier in business development based on mobile positioning data is the legal/data protection situation. This is also the main reason why it was not possible to acquire pilot data that could be used in the scope of this feasibility study.

On the other hand the lessons learnt in trying to access mobile positioning data in Germany, give important insights for future plans to work with this data inside and outside Germany:

- Contact with the right persons within all relevant MNOs (in the light of the merger of O2 and E-Plus) has been established.
- The MNOs are now aware of the general interests of tourism (statistics) for their future services.
- There is now a clear picture of the possibilities/data quality under the present legal framework.
- Data quality would require a change of legislation in Germany and substantial investments to get the data processing up and running (as in Estonia).
- Tourism indicators outside the Regulation 692/2011 could probably be more easily implemented: e.g. measure event/attraction tourism (see Report 4 Section 3.3.2.).

A general alternative put forward by MNOs and Data Protection Authority is the use of a large sample of Opt-In contacts to measure tourism.

6.5. Feasibility of Data Access in Other Countries

A total of 76 MNOs from Europe were contacted via different channels in order to introduce the current study and to discuss the possibilities of obtaining the pilot data. By the time the current report was finalised, there is a pending data from all Finnish MNOs (waiting for the approval of data protection agency for MNOs to be able to process the data internally and provide aggregated results to the consortium), German MNOs have declined the request for specifically usable (longitudinal) data and suggested using the alternative data of temporarily limited subscriber identification for financial compensation, which is not possible within the current project's scope. French MNOs can provide aggregated data on regional level, but have requested financial compensation as this is their business product.

A response from one Belgian MNO is still expected that considers providing the data for the project possibly free of charge. As Dutch Central Bureau of Statistics (CBS) have been able to work with the data of one Dutch MNO, the request has been submitted concerning the possibility to use the data or results in the current study, but the legal agreement for such use is still pending (from the MNO). However the data used by CBS is based on aggregated raw results, which is probably not usable for proposed methodology.

The survey discovered very few statistical offices and municipalities using such data. Even if the organisations are interested in the usage of mobile positioning data, they have not got access to it yet.

Although Statistical Office of Slovenia is starting the project with at least one Slovenian MNO in January 2014, it is still in the middle of the legal discussion concerning the applicability of the national statistics act to gain the right to access MNOs databases. The main concentration is on domestic data, but technical specifications are not yet established. Negotiations by the consortium with one other Slovenian MNO are still on-going if it would be possible to use their data in the project.

The consortium has been in contact with the Central Statistics Office Ireland (CSO) who have been working towards the possibility of using mobile data from local MNOs. Unfortunately by the time of the current report, the CSO has not been able to proceed with all regulatory procedures. The consortium has also been in direct contact with Irish MNOs and although they have been optimistic, they have expressed the neutral position until the CSO have cleared the legal procedures and there is an official legal basis for providing the data for CSO and possibly the current project.

Austrian MNOs have been contacted already prior to the current project with the hope of analysing the data with Austrian Statistical Office. Some portion of the data was received but the MNO has discontinued their efforts on the use of the data at this moment. In addition, the data already received is not usable for tourism statistics.

The Czech Tourist Authority has been able to work with a portion of mobile data concerning the analysis of the visitors to different events and locations. They were kind enough to introduce the specifics of the data. However, this data is not usable to assess the methodology discussed in this project.

The consortium has contacted jetsetme.com, a project by Telefonica/O2 concerning the roaming data of all MNOs within the Telefonica/O2 group. JetSetMe is a visualisation of the almost real-time roaming habits of mobile phones using SIM cards on O2's network while roaming in Europe. The consortium was refused the access or assessment of the data used by jetsetme.com within the time scope of the project.

7. Conclusion

The objective of this report was to make a thorough analysis of different aspects of accessibility to the mobile positioning data in EU. This includes the description of and discussion on regulatory, privacy protection, technological, financial and to some extent methodological barriers as well as discussion on possible solutions where possible.

For the research for the report, an online survey and interviews with experts and stakeholders were conducted; legal analysis was made on the regulatory situation on EU and four Member States (Estonia, Finland, France, Germany); efforts to obtain pilot data from MNOs in corresponding countries as well as other countries were undertaken; analysis of the different technological, methodological and financial aspects were done. A thorough description of the technical aspects of the access to the data has been presented.

The survey showed that many organisations are interested in using mobile positioning data for the production of statistics and research. NSIs anticipate the value that this data source could bring to their field and there are some that have tried reaching out to operators, with only a few gaining access to the data as of yet. The main constraints to access are regulatory, privacy-related, and financial.

What interviews clearly brought out is that MNOs have the most concern regarding regulations. The value in providing data is understandable, and most MNOs sympathise with

the idea of using the data for statistics, but clearly state that many issues need to be dealt with before that is possible. MNOs are looking into utilizing their data commercially and are interested in tourism statistics as one potential domain. However financial concerns, possible disclosure of business secrets and the effect of public opinion are also considered important besides the regulatory concerns. From the privacy protection point of view, it is considered important that the sensitive private data of subscribers is used according to the legislation and is also presented to the public as appropriate, lawful and that the objective is not to track individual people. The MNOs are confident the technology is mature enough not to pose problems and the main technical problem is making sure the data is handled in a way that is methodologically correct.

Legal analysis concludes that there is no single clear understanding at the moment on how mobile data can be used in generating statistics. Although the EU regulations and directives are same for every Member State, the underlying national regulations are implemented differently and do not propose a single simple way for NSIs to obtain such data. Although in some Member States it is possible to implement the effective national statistics act in order to enforce the requisition of the data from MNOs, in most Member States the act has to be amended in order for the statistics authority to have specific obligation to collect data from MNOs for specific purpose. Unfortunately, as legislation in this part differs, there is no single clear suggestion on how to do it conformably in every Member State. The same applies to the specific procedure and processes on the data acquisition – it is not clearly stated if the transmitted data should be personal or anonymous, processed or simply aggregated (though not essential, transmitting personal data provides the best outcomes in terms of methodology). This is also important from the point of view of the methodological harmony between Member States as implementation of the common methodology and processing logic would result in the smallest differences in the statistical results between Member States.

The alternative option is to introduce mobile data as required data at European legislation level (i.e. Regulation 692/2011) that is directly applicable in all Member States. However, for both options (local and EU level) the prerequisite is that mobile data is necessary for the performance of a specific task (e.g. the question becomes whether to meet the prerequisite the data should be inevitably necessary for the performance of the task or is it sufficient if the performing of a task is easier, more efficient, etc. as a result of processing the data compared to possible alternative measures). Such prerequisite is discussed in the following reports of the study.

The comprehensive description of initial extraction of the data from MNOs provides readers with opportunity to go step by step through the essential processes of the extraction of the data. These processes are often not simple and require sophisticated system to be able to extract and prepare valuable data for tourism statistics. The process is tightly connected with the methodological steps described in the Report 3a that continues the description of the processes used in frame formation and data compilation.

MNOs are big businesses with considerable revenues and with highly complex technological systems. Financial and business aspects covered in this report present an insight to the MNOs concerns towards the use of the data. The cost of the implementation of the technology to extract data is very high and has to be taken into account. The presented figures of the implementation are rough indications because the actual costs of either the pilot data, implementation of the automated data extraction and processing system, and the maintenance of the system can only be calculated individually by MNOs. Those costs might include the potential licencing of the system as well as patent licencing on specific methodologies. Apart from the financial burden, MNOs often have to consider the reaction from the publicity as mobile data is highly sensitive and private data. Negative public opinion can result in a loss of customers that can be much more damaging than implementation or maintenance cost for the MNOs. Still, MNOs see a potential in utilising their data in various fields, so the interest from MNOs is high. This benefits the progress also for statistics authorities as MNOs might be willing to 'experiment' with the data for the purpose of discovering new possibilities and potentially new revenue sources for their business.

During the project several MNOs were contacted for interviews, information on accessibility and technical consultation. MNOs were also asked if they could provide pilot data for the current study. Pilot data from three countries was accessed, but only the Estonian data proved to be of a sufficient quality for the required empirical tests in this study. Although contacted MNOs were interested in the project and showed interest in the outcome, only few responded to the possibility of providing the data. The main obstacles were connected to the regulations and missing legal basis for providing the data. The data that the consortium could get access to (outside of Estonia) was not suitable for the calculations and coherence analysis in Reports 3a and 3b. With many potential data providers, the process of legal clarification took too much time (is still on-going) and therefore even if the data access is granted, it will not be possible to analyse the data within the remaining time of the project.

The consortium members would like to thank the contributors to the current study, the respondents of the survey and interviews, contacted MNOs and legal experts.

References

Legal documents

Act on Legal Obligation, Coordination and Confidentiality in the Field of Statistics 51-711/1951: http://www.insee.fr/en/insee-statistique-publique/qualite/Loi_n_51-711_du_7_juin_1951_version_consolidee_30_juin_2010_EN.pdf

Act on the Openness of Government Activities 621/1999: http://www.finlex.fi/en/laki/kaannokset/1999/en19990621

Act on the Protection of Privacy in Electronic Communications 516/2004: http://www.finlex.fi/en/laki/kaannokset/2004/en20040516

Communications Market Act 393/2003: http://www.finlex.fi/en/laki/kaannokset/2003/en20030393

Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF

Directive 2002/58/EC (as amended by 2009/136/EC) of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:EN:PDF

Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF

Electronic Communications Act 2004: http://www.legaltext.ee/text/en/X90001K4.htm

European Patent Convention (EPC): http://www.epo.org/law-practice/legal-texts/epc.html

Federal Data Protection Act (BDSG) 2003: http://www.bfdi.bund.de/EN/DataProtectionActs/Artikel/BDSG_idFv01092009.pdf?__blob=publicationFile

Federal Statistics Act (BStatG): https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/010a_BStatG_Engl.pdf

General Data Protection Regulation [Proposed regulation]: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0009:FIN:EN:PDF

Personal Data Protection Act 2003: http://www.legaltext.ee/text/en/X70030.htm

Gesellschaft für Antriebstechnik mbH & Co. KG (GAT) v Lamellen und Kupplungsbau Beteiligungs KG. C-4/03 (2006) Judgment of the European Court of Justice (First Chamber) of 13 July 2006: http://curia.europa.eu/juris/liste.jsf?language=en&num=C-4/03

Heritage Code 2004: http://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006074236

Information Technology, Data Files and Civil Liberties Act 78-17/1978: http://www.cnil.fr/fileadmin/documents/en/Act78-17VA.pdf

Official Statistics Act 2010: https://www.riigiteataja.ee/en/eli/511112013008

Personal Data Act 523/1999: http://www.finlex.fi/en/laki/kaannokset/1999/19990523

Postal and Electronic Communications Code: http://www.arcep.fr/fileadmin/reprise/textes/lois/cpce-decrets.pdf

Regulation (EU) No 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 (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the statistical Programmes of the European Communities. Link: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:en:PDF

Roche Nederland BV and Others v Frederick Primus and Milton Goldenberg. C-539/03 (2006) Judgment of the European Court of Justice (First Chamber) of 13 July 2006: http://curia.europa.eu/juris/liste.jsf?language=en&num=C-539/03

Telecommunications Act (TKG) 2004: http://www.bfdi.bund.de/EN/DataProtectionActs/Artikel/TelecommunicationsAct-TKG.pdf

Telemedia Act (TMG): http://www.cgerli.org/fileadmin/user_upload/interne_Dokumente/Legislation/Telemedia_Act__TMA_.pdf

The Statistics Act 280/2004: http://tilastokeskus.fi/meta/lait/2013-09-02_tilastolaki_en.pdf

Publications

A1 Traffic Data Stream: Movement Data in Mobile Telephone Network as Data Source for Marketing, Research and Planning. Vienna, 17 December 2009 [Press Release]: http://www.wigeogis.com/en/pdf/news/NEWS17122009.pdf

Accenture (2013) The Future Communications Service Provider: http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-The-Future-CSP-Converged-Digital-World.pdf Ahas, R. and Mark, Ü. (2005) Location based services—new challenges for planning and public administration? *Futures*, 37(6), pp 547–561

Appelt, C.W. (2007, November 11) Enforcement of Patents in Europe – Germany as an Example. Ip4inno [Weblog]: http://www.ip4inno.eu/index.php?id=187

Apple Inc. (2011, April 27) Apple Q&A on Location Data [Press release]: http://www.apple.com/pr/library/2011/04/27Apple-Q-A-on-Location-Data.html

Article 29 Data Protection Working Party Opinions: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/index_en.htm

ATKearney (2012, March) Mobile Network Operators as Smart Enablers. Ideas and Insights: http://www.atkearney.com/paper/-/asset_publisher/dVxv4Hz2h8bS/content/mobile-network-operators-as-smart-enablers/10192

Balbi, A. (2006, July 19) Il Grande Fratello del traffico così si controlla la mobilità. *La Repubblica*: http://www.repubblica.it/2006/07/sezioni/cronaca/cellulari-traffico/cellulari-traffico/cellulari-traffico.html

Bauman, Zygmunt (2013) *Liquid Fear*. John Wiley & Sons, 200p

Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Medicine*, Vol 8 No 8.

Biermann, K. (2011, March 26) Data Protection: Betrayed by our own data, Zeit Online: http://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz

Big Data: Der vermessene Mensch (2013) *GEO*, 8/2013

Cardon, Dominique (2010, December 3) 'En finir avec le culte du secret et de la raison d'Etat'. *Le Monde*: http://www.lemonde.fr/retrospective/article/2010/12/03/en-finir-avec-le-culte-du-secret-et-de-la-raison-d-etat_1448555_1453557.html

Chen, B.X. (2011, April 21) Why and How Apple Is Collecting Your iPhone Location Data. *Wired*: http://www.wired.com/gadgetlab/2011/04/apple-iphone-tracking/

Cheng, J. (2009, August 25) Arterial, crowdsourced traffic info comes to Google Maps. *Ars Technica* [Weblog]: http://arstechnica.com/tech-policy/2009/08/arterial-crowdsourced-traffic-info-comes-to-google-maps/

Cisco (2013, February) The Future of Mobile Networks: http://www.cisco.com/web/about/ac79/docs/sp/Future-of-Mobile-Networks.pdf

Coleman Parkes (2013) Amdocs Survey: Consumers Will Share Personal Data… at a Price. Amdocs: http://www.amdocs.com/News/Pages/amdocs-personal-data-consumer-survey.aspx

Forsa Institute (2008) Forsa-Umfrage: Vorratsdatenspeicherung verhindert sensible Gespräche. At the request of Working Group on Data Retention, eco - Association of the German Internet Industry, German Association for Specialized Journalists and JonDos GmbH: http://www.daten-speicherung.de/data/forsa_2008-06-03.pdf

Foucault, M. (1995) *Discipline and Punishment*. Vintage Books, New York

Google (2009, August 25) The bright side of sitting in traffic: Crowdsourcing road congestion data. Google Official Blog [Weblog]: http://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html

Government Study Identifies 52 'Population Centers' (2013, May 15) *Estonian Public Broadcasting*: http://news.err.ee/Politics/134b3355-49ba-4eda-9010-a2e1f4986ae7

Guide on Security of Personal Data (2010) CNIL: http://www.cnil.fr/fileadmin/documents/en/Guide_Security_of_Personal_Data-2010.pdf

Hypponen, M. (2011, April 21) Actually, iPhone Sends Your Location to Apple Twice a Day, *F-Secure* [Weblog]: http://www.f-secure.com/weblog/archives/00002145.html

Infas (2010) Der überwachte Bürger zwischen Apathie und Protest - Erste Ergebnisse. At the request of Institut für Sicherheits- und Präventionsforschung e.V. (ISIP): http://www.vorratsdatenspeicherung.de/images/infas-umfrage.pdf

Koskela, H. (2009) Fear and Its Others. *The SAGE Handbook of Social Geographies*. SAGE, 2009

KPMG (2013) The future of mobile networks: http://www.kpmg.com/global/en/issuesandinsights/articlespublications/mobile-evolution/pages/future-mobile-networks.aspx

Leben nach Zahlen: Big Data: Wie Staaten und Konzerne berechnen, was wir tun werden (2013). *Der Spiegel*, 20/2013, pp 1-154.

MacManus, R. (2009, April 5) Real Time Cities, or Just Info Porn? ReadWriteWeb [Weblog]: http://readwrite.com/2009/04/05/real_time_cities_or_info_porn

Macura, B. (2011, April 8) Das Geschäft mit den Verkehrsdaten. *ÖRF*: http://help.orf.at/stories/1682467/

Maras, M.-H. (2012) The social consequences of a mass surveillance measure: What happens when we become the 'others'? *International Journal of Law, Crime and Justice*, Vol 40 No 2, pp 65–81

Massachusetts Institute Of Technology (2007, September 17). 'Wiki City Rome' To Draw A Map Like No Other. *ScienceDaily*: http://www.sciencedaily.com/releases/2007/09/070907104822.htm

McNeil, D.G. (2011, September 5) Haiti: Cellphone Tracking Helps Groups Set Up More Effective Aid Distribution, Study Says. *The New York Times*: http://www.nytimes.com/2011/09/06/health/06global.html

Milberg, Sandra J., Smith, H. Jeff, and Burke, Sandra J. (2000) Information privacy: Corporate management and national regulation. *Organization Science*, Vol 11 No 1, pp 35–57

Mobile phones help to target disaster aid, says study. *BBC*, 2 Sept 2011: http://www.bbc.co.uk/news/technology-14761144

mobilkom austria (2003) A1 VERKEHR bringt Stau- sowie Radarinfos auf das Handy und bietet Online-Navigationshilfe [Press Release]: http://www.pressetext.com/news/20030617010

Moses, A. (2011, May 6) Outrage over TomTom speed traps for motorists. *The Sunday Morning Herald*: http://www.smh.com.au/digital-life/cartech/outrage-over-tomtom-speed-traps-for-motorists-20110506-1ebc2.html

Nissenbaum, H.F. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto: Stanford University Press.

Oliver, N. (2013, September 5) Combating global epidemics with big mobile data. *The Guardian*: http://www.theguardian.com/media-network/media-network-blog/2013/sep/05/combating-epidemics-big-mobile-data

Page, Benjamin I., Shapiro, Robert Y. and Dempsey, Glenn R. (1987) 'What Moves Public Opinion?' *The American Political Science Review*, Vol 81 No 1, pp 23-44

Pew Research Center (2012) Privacy and Data Management on Mobile Devices: http://pewinternet.org/~/media//Files/Reports/2012/PIP_MobilePrivacyManagement.pdf

Pew Research Center (2013) Location-based Services: http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_Location-based%20services%202013.pdf

PwC (2013) The Global State of Information Security® Survey 2014: http://www.pwc.com/gx/en/consulting-services/information-security-survey/giss.jhtml#

Rajalo, S. (2013, June 14) Ärileht esitleb: mobiili asukohaandmete kasutamisest saab tulevikuäri. Ärileht: http://arileht.delfi.ee/news/tarbija/arileht-esitleb-mobiili-asukohaandmete-kasutamisest-saab-tulevikuari.d?id=66285950

Ratti, C., Shevtsuk, A. et al (2005) Mobile Landscapes: Graz in Real Time, conducted by MIT & Ratti Associates: http://senseable.mit.edu/graz/

Richmond, S. (2011, April 21) Apple promises fix for iPhone tracking 'bug'. *The Telegraph*: http://www.telegraph.co.uk/technology/apple/8478586/Apple-promises-fix-for-iPhone-tracking-bug.html

Rothkopf, D. (2013, October 29) Mr. President, We Can Handle the Truth: Why it's time for the White House to get ahead of the NSA scandal. *Foreign Policy*: http://www.foreignpolicy.com/articles/2013/10/29/mr_president_we_can_handle_the_truth_white_house_nsa_scandal

Sheehan, K. B. (2002). Toward a typology of Internet users and online privacy concerns. *The Information Society*, Vol 18 No 1, pp 21-32.

Smith, H.J.; Milberg, J.S.; and Burke, J.S. (1996) Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly*, Vol 20 No 2, pp 167–196

Son, J.-Y., and Kim, S.S. (2008) Internet users' information privacy-protective responses: a taxonomy and a nomological model. *MIS Quarterly*, Vol 32 No 3, pp 503–529.

Sorrel, C. (2009, August 26) Google Mobile Maps To Crowd-Source Traffic Data. *Wired* [Weblog]: http://www.wired.com/gadgetlab/2009/08/google-mobile-maps-to-crowd-source-traffic-data/

Special Eurobarometer 359 (2011) Attitudes on Data Protection and Electronic Identity in the European Union. TNS Opinion & Social at the request of Directorate-General Justice, Information Society & Media and Joint Research Centre. http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf

Tata Consultancy Services (2013) The Emerging Big Returns on Big Data: http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf

Telefonica (2012, November 14) Telefonica Dynamic Insights launches 'Smart Steps' in the UK. Telefonica Digital Hub [Weblog]: http://blog.digital.telefonica.com/?press-release=telefonica-dynamic-insights-launches-smart-steps-in-the-uk

Telefonica (2013) Big data and social good: Nuria Oliver. Telefonica Dynamic Insights [Weblog]: http://dynamicinsights.telefonica.com/1146/big-data-and-social-good-nuria-oliver

Telefonica hopes 'big data' arm will revive fortunes (2012, October 9) BBC: http://www.bbc.co.uk/news/technology-19882647

Tofel, K.C. (2009, August 25) Crowdsourcing Brings Better Traffic Data to Google Maps, Even Without iPhone Data. *GigaOm* [Weblog]: http://gigaom.com/2009/08/25/crowdsourcing-brings-better-traffic-data-to-google-maps-except-on-iphone/

TomTom's CEO Harold Goddijn on Data Privacy, 27 April 2011 [Video]: http://www.youtube.com/watch?v=Zc_cGepf1qg

Waters, R. (2006, September 1) Rome - like you've never seen it before. *Financial Times*: http://www.ft.com/intl/cms/s/0/6a4c3c82-3956-11db-a21d-0000779e2340.html

Watson, S.M. (2011, October 10) The Latest Smartphones Could Turn Us All Into Activity Trackers. *Wired*: http://www.wired.com/opinion/2013/10/the-trojan-horse-of-the-latest-iphone-with-the-m7-coprocessor-we-all-become-qs-activity-trackers/

Williams, C. (2011, April 21) Apple under pressure over iPhone location tracking. *The Telegraph*: http://www.telegraph.co.uk/technology/apple/8466357/Apple-under-pressure-over-iPhone-location-tracking.html

Williams, C. (2011, April 28) Police use TomTom data to target speed traps. *The Telegraph*: http://www.telegraph.co.uk/technology/news/8480195/Police-use-TomTom-data-to-target-speed-traps.html

Websites

Patent Cooperation Treaty (PCT): http://www.wipo.int/pct/en/

PATENTSCOPE [Database]: http://patentscope.wipo.int/

World Bank [website] http://databank.worldbank.org/

World Intellectual Property Organization (WIPO): http://www.wipo.int/

Annex 1. List of Technical Abbreviations

Abis – The interface (communication protocol) between the BTS (antenna) and BSC (antenna controller).

API – Application Programming Interface

AuC – Authentication centre

BSC – Base Station Controller (2G calls, messaging, internet)

BSS – Base Station Subsystem

BTS – Base Transceiver Station, 2G antenna

CC – Country Code (from MSISDN)

CDR – Call Detail Records

CGI – Cell Global Identity

DDR – Data Detail Records (same as IPDR)

eNodeB – 4G antenna

GGSN – Gateway GPRS Support Node (2G, 3G internet)

GPRS – General Packet Radio Service

HLR – Home Location Register

HSS – Home Subscriber Server

IMSI – International Mobile Subscriber Identity

IPDR – Internet Protocol Data Records (same as DDR)

LA – Location Area

LBS – Location-based Services

LTE – Long Term Evolution, 4G LTE, is a standard for wireless communication of high-speed data for mobile phones and data terminals

M2M – Machine-to-Machine (SIMs)

MCC – Mobile Country Code (from IMSI)

MD – Mobile Device, mobile phone

MME – Mobility Management Entity (4G calls, messaging)

MMS – Multimedia Message Service

MNC – Mobile Network Code

MNO – Mobile Network Operator

MPS – Mobile Positioning System

MSC – Mobile Switching Centre

MSIN – Mobile Subscription Identification Number

MSISDN – Mobile Station Integrated Services Digital Network

NDC – National Destination Code

NMS – Network Management System

NodeB – 3G antenna

NPA – Number Planning Area

NSS – Network Subsystem

OSS – Operation and Support Subsystem

PLMN – Public Land Mobile Network. A single MNO network within a country

PSTN – Public Switched Telephone Network

RNC – Radio Network Controller (3G calls, messaging, internet)

SGSN – Service GPRS Support Node (for Internet traffic 2G, 3G)

SGW – Serving Gateway

SIM – Subscriber Identity Module, an integrated circuit that securely stores the International Mobile Subscriber Identity (IMSI) and the related key used to identify and authenticate subscribers on mobile telephony devices

SMS – Short Message Service

SN – Subscriber's number

TMSI – Temporary Mobile Subscriber Identity

UDR – Usage Detail Records

VLR – Visitor Location Register

Annex 2. EC Regulatory Documents

- Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/directive-1995_46_ec.pdf
  Annexed file: Directive 1995_46_EC.pdf
- Directive 2002/58/EC (as amended by 2009/136/EC) of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector.
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:EN:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/directive-2002_58_ec.pdf
  Annexed file: Directive 2002_58_EC.pdf
- Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/directive-2006_24_ec.pdf
  Annexed file: Directive 2006_24_EC.pdf
- Regulation (EU) No 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.
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/regulation-2011_692_eu.pdf
  Annexed file: Regulation 2011_692_EU.pdf

- Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the statistical Programmes of the European Communities.
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:en:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/regulation-2009_223_ec.pdf
  Annexed file: Regulation 2009_223_EC.pdf
- Article 29 Data Protection Working Party Opinion.
  Weblink (official EU legislation URL): http://ec.europa.eu/justice/data-protection/article-29/index_en.htm
- General Data Protection Regulation (Proposed regulation).
  Weblink (official EU legislation URL): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0009:FIN:EN:PDF
  Weblink (project URL): http://mobfs.positium.ee/data/uploads/task-2/general-data-protection-regulation-proposed-regulation.pdf
  Annexed file: General Data Protection Regulation (Proposed regulation).pdf

Annex 3. Memorandum on Legal Regulation on the Use of Mobile Positioning Data in the European Union and Estonia, Finland, Germany and France

Legal memorandum by Borenius can be found at the following project URL:
http://mobfs.positium.ee/data/uploads/task-2/legal-memorandum.pdf

Annex 4. Online Survey Form

The online survey can be found at the following URL:

https://docs.google.com/forms/d/15wRJ4obYBYa6uH9q7hJzK5inoGci_51rg87cguyisUM/viewform

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

Dear Mr./Mrs./Ms.,

This questionnaire is a part of the Eurostat (European Commission) project titled, "Feasibility study on the use of mobile positioning data for tourism statistics" where the key objective is to investigate the possibilities of use of anonymous aggregated data from mobile network operators in generating tourism statistics.

The aim of these questions is to understand the aspects of accessibility barriers and opportunities of mobile positioning data in the tourism sphere as well as gain the experience of organisations like yours.

This project is conducted by a consortium composed of Positium (Estonia), Statistics Finland, Statistics Estonia, the French Institute of Science and Technology for Transport, Development and Networks, the Institute for Tourism Research in Northern Europe (Germany) and the University of Tartu.

The results of this project will be publicly available by Eurostat in the summer 2014. If you want to receive more information on the project or on using anonymous mobile positioning data, please let us know directly.

If you feel that this questionnaire does not concern you and/or you know another person in your organisation who can answer there questions better, please refer this link to him/her with appropriate questionnaire ID.

Consortium member responsible for the questionnaire (Laura Altin – laura.altin@positium.ee).

==========

Mobile positioning data in this project refer to the large-scale location data of subscribers of mobile network operators that are processed and stored in operators' systems. This is highly sensitive but also very valuable data that could be used anonymously and aggregated thus ensuring the privacy of the subscribers and providing valuable insights in fields like tourism statistics (the subject of this project) and many other domains (transportation and urban planning, regional development, geomarketing, safety and security, crisis management, research, etc.).

The views expressed in this study do not necessarily reflect the official position of the European Commission or Eurostat.

The questionnaire will be analysed anonymously and the answers will not be connected to specific respondent in the analysis and reports. Questionnaire ID / contact information will only be used to contact the respondent in case of any further questions.

* Required

1. **Plese enter Your questionnaire ID!**
   You can find this in the e-mail about the questionnaire.

   .....

2. **Your name (optional)**
   If You do not have the questionnaire ID, please let us know who You are!

   .....

3. **Your organisation (optional)**
   If You do not have the questionnaire ID, please let us know who You are!

   .....

4. **Your e-mail (optional)**
   If You do not have the questionnaire ID, please let us know who You are!

   .....

5. **What organization do You represent?** *
   Mark only one oval.

   ◯ Mobile Network Operator *After the last question in this section, skip to question 7.*
   ◯ National (Regional) Statistical Office *After the last question in this section, skip to question 45.*
   ◯ National Data Protection Agency *After the last question in this section, skip to question 35.*
   ◯ Central Bank *After the last question in this section, skip to question 45.*
   ◯ Research Institute *After the last question in this section, skip to question 45.*
   ◯ Municipality, Regional government *After the last question in this section, skip to question 45.*
   ◯ National or regional tourism organisation *After the last question in this section, skip to question 45.*
   ◯ Public & government Organizations (ministries, departments) *After the last question in this section, skip to question 45.*
   ◯ Eurostat directorate / European Commission *After the last question in this section, skip to question 45.*
   ◯ Private enterprise *After the last question in this section, skip to question 45.*
   ◯ Other: ..... *After the last question in this section, skip to question 45.*

6. **Comments**

   .....
   .....
   .....
   .....
   .....

A1

7. **Have You been approached by anybody for the purpose of using anonymous mobile positioning data of Your subscribers?** *
   Mark only one oval.
   - ◯ Yes
   - ◯ No
   - ◯ Not aware of

8. **Are You providing / have You been providing the anonymous mobile positioning data to anybody?** *
   Mark only one oval.
   - ◯ Yes *After the last question in this section, skip to question 11.*
   - ◯ No *After the last question in this section, skip to question 24.*
   - ◯ Not aware of *After the last question in this section, skip to question 24.*

9. **If You are not providing the data, but there is interest in that, what are the reasons for declining?**

   .....
   .....
   .....
   .....
   .....

10. **Comments**

    .....
    .....
    .....
    .....
    .....

A2

11. **What was the reason You provided the data?** *

    .....
    .....
    .....
    .....
    .....

12. **Where is/was this data used?** *
Check all that apply.

- [ ] Official government statistics
- [ ] Research
- [ ] Business use of the data
- [ ] Other: .....

13. **In what fields the data are used / have been used?** *
Check all that apply.

- [ ] Tourism statistics
- [ ] Transportation planning
- [ ] Geomarketing
- [ ] Urban planning
- [ ] Migration research
- [ ] Commuter studies
- [ ] Crisis management
- [ ] Traffic monitoring
- [ ] Other: .....

14. **What kind of data are/were You providing?** *
Check all that apply.

- [ ] Inbound roaming data
- [ ] Outbound roaming data
- [ ] Domestic subscribers data
- [ ] CDR (Call Detail Records)
- [ ] DDR (Data Detail Records)
- [ ] Unspecified network activity data
- [ ] More detailed data (probes, MSC-level, location updates etc.)
- [ ] Other: .....

15. **What was the form that You provided the data?** *
Mark only one oval.

- ◯ Aggregated results
- ◯ Raw data, no anonymization or aggregation
- ◯ Raw anonymous data
- ◯ Other: .....

16. **What is Your model for providing the data?** *
Check all that apply.

- ☐ Directly to users (sell)
- ☐ Licence the data to brokerage partner
- ☐ Provide data for government based on request for official use
- ☐ Other: .....

17. **Who are using this data?**
If possible, list the users of Your data with contacts so we could contact them with this questionnaire!

.....

.....

.....

.....

.....

18. **How is/was the data provided?** *
Check all that apply.

- ☐ One-time project
- ☐ Provide data continuously
- ☐ Other: .....

19. **Can You describe the specifications of the data? Format? Amount? Time-period? Number of events? etc.**

.....

.....

.....

.....

.....

20. **Information on the usage of the data**
If there is any kind of additional information, documentation, methodology description, web-page etc. that You could provide us?

.....

.....

.....

.....

.....

21. **What kind of experience have You had concerning public opinion on the use of mobile positioning data?** *

.....

.....

.....

.....

.....

22. **Do You know if other operators are providing the data in same or other projects?** *
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Not aware of

23. **Comments**

.....

.....

.....

.....

.....

A3

24. **Do You use any kind of mobile positioning data in Your internal analyses?** *
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Not aware of

25. **If Yes, please describe where is the data used and for what purposes?**

.....

.....

.....

.....

.....

26. **Comments**

.....

.....

.....

.....

.....

A4

27. **In Your opinion, what are the main obstacles, risks for providing anonymous aggregated mobile positioning data?** *

Check all that apply.

- [ ] Privacy concerns (bad publicity)
- [ ] Regulatory and legislation obstacles
- [ ] Exposure of business secrets
- [ ] High implementation and maintenance cost
- [ ] No business value
- [ ] Technological issues
- [ ] Other: .....

28. **What regulatory acts, directives control the use of mobile positioning data?**

EU and national level legislation

.....

.....

.....

.....

.....

29. **Do You have to consult or get permission from Data Protection Agency and/or undergo any legal procedures concerning the use of the data?** *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Not aware of

30. **What are the business secrets associated with mobile positioning data?**

31. **Are there any internal rules or policy on using and providing mobile positioning data?** *
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Not aware of

32. **If Yes, can You describe shortly what are they and who controls the use of the data?**

33. **What is Your opinion on using mobile positioning data? Do You see any future for that? Is there a business for You?**

34. **Comments**

B1

Skip to question 72.

35. **Are You aware of any use cases in Your country, where mobile positioning data were used?** *
Mark only one oval.

- ◯ Yes
- ◯ No

36. **Has Your organisation been dealing with the cases of using mobile positioning data?** *
Mark only one oval.

- ◯ Yes *After the last question in this section, skip to question 41.*
- ◯ No *After the last question in this section, skip to question 72.*
- ◯ Not aware of *After the last question in this section, skip to question 72.*

37. **In Your opinion, what are the main obstacles, risks for providing anonymous aggregated mobile positioning data?** *
Check all that apply.

- ☐ Privacy concerns
- ☐ Regulatory and legislation obstacles
- ☐ Exposure of business secrets
- ☐ High implementation and maintenance cost
- ☐ No business value
- ☐ Technological issues
- ☐ Other: .....

38. **What regulatory acts, directives control the use of mobile positioning data?** *
EU and national level legislation

.....

.....

.....

.....

.....

39. **If mobile network operators want to use anonymous aggregated mobile positioning data, what legal processes do they have to proceed?** *

.....

.....

.....

.....

.....

40. **Comments**

.....

.....

.....

.....

.....

B2

Skip to question 72.

41. **Where is/was this data used?** *

Check all that apply.

- [ ] Official government statistics
- [ ] Research
- [ ] Business use of the data
- [ ] Mobile network operator internal use
- [ ] Other: .....

42. **In what fields the data are used / have been used?** *

Check all that apply.

- [ ] Tourism statistics
- [ ] Transportation planning
- [ ] Geomarketing
- [ ] Urban planning
- [ ] Migration research
- [ ] Commuter studies
- [ ] Crisis management
- [ ] Traffic monitoring
- [ ] Other: .....

43. **Please describe how was Your organisation involved in this process?** *

.....

.....

.....

.....

.....

44. **Comments**

.....

.....

.....

.....

.....

C1

Skip to question 72.

45. **Does Your organisation use / generate any kind of tourism statistics?** *
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Not aware of

46. **Are You aware of the possibilities of using mobile positioning data?** *
Mark only one oval.

- ◯ Yes
- ◯ No

47. **Does Your organisation use mobile positioning data?** *
Mark only one oval.

- ◯ Yes *After the last question in this section, skip to question 49.*
- ◯ No *After the last question in this section, skip to question 64.*
- ◯ Not aware of *After the last question in this section, skip to question 64.*

48. **Comments**

.....

.....

.....

.....

.....

C2

49. **Where is this data used?** *
Check all that apply.

- [ ] Official government statistics
- [ ] Research
- [ ] Business use of the data
- [ ] Other: .....

50. **In what fields the data are used / have been used?** *
Check all that apply.

- [ ] Tourism statistics
- [ ] Transportation planning
- [ ] Geomarketing
- [ ] Urban planning
- [ ] Migration research
- [ ] Commuter studies
- [ ] Crisis management
- [ ] Traffic monitoring
- [ ] Other: .....

51. **How did You get access to the data?** *
Mark only one oval.

- ◯ Directly from mobile network operator(s)
- ◯ From mobile positioning data providers (other than MNO)
- ◯ From research institutes
- ◯ Other: .....

52. **Are You allowed to use the name(s) of the mobile network operator(s) whose data You are using?** *
Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Some yes, some no
- ◯ Don't know (if You get data from provider)

53. **Who do You get this data from?**
If possible, name the mobile network operator(s) name(s)! If not/don't know, name the specific provider of the data!

.....

.....

.....

.....

.....

54. **What kind of data is used?** *
Check all that apply.

- [ ] Inbound roaming data
- [ ] Outbound roaming data
- [ ] Domestic subscribers data
- [ ] CDR (Call Detail Records)
- [ ] DDR (Data Detail Records)
- [ ] Unspecified network activity data
- [ ] More detailed data (probes, MSC-level, location updates etc.)
- [ ] Other: .....

55. **What is the form of the data?** *
Mark only one oval.

- ◯ Aggregated results
- ◯ Raw data, no anonymization or aggregation
- ◯ Raw anonymous data
- ◯ Other: .....

56. **How was the data provided?** *
Check all that apply.

- [ ] One-time project
- [ ] Provide data continuously
- [ ] Other: .....

57. **How many mobile network operators are providing the data?** *
Mark only one oval.

- ◯ One
- ◯ Two or more but not all
- ◯ All
- ◯ Other: .....

58. **What procedures did You have to go through to get the data?** *

.....

.....

.....

.....

.....

59. **Can You describe the specifications of the data? Format? Amount? Time-period? Number of events? etc.**

.....

.....

.....

.....

.....

60. **What was the reason the provider did not decline You the data?**
Why did they say yes and provided You with the data?

.....

.....

.....

.....

.....

61. **What problems does the mobile positioning data solve in Your organisation? What value does it bring?** *

.....

.....

.....

.....

.....

62. **Information on the usage of the data**
If there is any kind of additional information, documentation, methodology description, web-page etc. that You could provide us?

.....

.....

.....

.....

.....

63. **Comments**

.....

.....

.....

.....

.....

C3

Skip to question 71.

64. **Are You interested in using mobile positioning data in Your work?** *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ No - mobile positioning data has no value in Your organisation's processes

65. **Have You approached mobile network operators or other organisations concerning obtaining mobile positioning data?** *

Mark only one oval.

- ◯ Yes *Skip to question 66.*
- ◯ No *Skip to question 71.*
- ◯ Not aware of *Skip to question 71.*

C4

66. **Who have You approached concerning getting mobile positioning data?** *

.....

.....

.....

.....

.....

67. **In what fields the data would be used?** *
Check all that apply.

- [ ] Tourism statistics
- [ ] Transportation planning
- [ ] Geomarketing
- [ ] Urban planning
- [ ] Migration research
- [ ] Commuter studies
- [ ] Crisis management
- [ ] Traffic monitoring
- [ ] Other: .....

68. **What problems would the mobile positioning data solve in Your organisation? What value would it bring?** *

.....

.....

.....

.....

.....

69. **What procedures did / do You have to go through?** *
Data Protection Agency, legislation, mobile network operator specific procedures?

.....

.....

.....

.....

.....

70. **What are the reasons You haven't get access to the data yet?** *

.....

.....

.....

.....

.....

C5

71. **In Your opinion, what are the main obstacles, risks for providing anonymous aggregated mobile positioning data?** *

Check all that apply.

- [ ] Privacy concerns
- [ ] Regulatory and legislation obstacles
- [ ] Exposure of business secrets
- [ ] High implementation and maintenance cost
- [ ] No business value
- [ ] Technological issues
- [ ] Other: _____

Final

72. **Anyone other You know, who has been working with data from mobile network operators?**

We would be happy to contact them! Please provide any contacts below!

_____

_____

_____

_____

_____

73. **Do You have any extra comments, suggestions, advice for use concerning this project and the use of mobile positioning data?**

_____

_____

_____

_____

_____

Annex 5. Expert Interview Guidelines

Mobile Network Operator (MNO)

This interview is a part of Eurostat's (European Commission) project "Feasibility study on the use of mobile positioning data for tourism statistics" where the key objective is to investigate the possibilities of the use of anonymous aggregated data from mobile network operators in generation of tourism statistics.

The aim of the interview is to understand the aspects of accessibility barriers and opportunities of mobile positioning data in tourism sphere as well as the experiences of organizations like yours.

The project is conducted by a consortium composed of Positium LBS (Estonia), Statistics Finland, Statistics Estonia, French Institute of Science and Technology for Transport, Development and Networks, The Institute for Tourism Research in Northern Europe (Germany) and University of Tartu.

The result of this project will be publicly available by Eurostat from the summer 2014. If you want to receive more information on the project or on using anonymous mobile positioning data, please let us know directly.

A) To your knowledge, are mobile positioning data already being used for one or more of following purposes:

1. … anonymized or non-anonymized identification or tracing of individual mobile devices for public or private purposes?

2. ... applications like traffic monitoring (FCD), police or emergency services etc?

3. … more specifically: e.g. for the enhancement of official statistics?

0

B) If yes: What do you know about these projects or applications? Please give more detailed information on each of the projects or applications (or name a reference):

1. Which companies/organizations were involved?

2. What is the description of the data that has been asked from you?

3. Have you provided them with the data?

4. What form/structured data have you provided?

5. If you have provided data, what was your reason?

6. If you don't provide mobile positioning data for everyone, what is the basis for your decision?

7. When providing the data, have you consulted or been into contact with Data Protection Agency and/or undergone any legal procedures concerning the use of the data?

8. Was/Is it a onetime project or a continuous operation? Is it still active?

9. What were the main learnings? Would you work with the same setup again? What would you change?

10. Is there any documentation on the project? Is it accessible?

C) How is such positioning data being used internally?

- Do you have any business interests in providing any kind of services based on your data?

D) What are, in your opinion, main barriers for the implementation of a positioning service for tourism statistics? Do you see any solutions? Please consider the following aspects.

1. Privacy and regulatory issues:

2. Financial and business issues:

3. Technological issues:

4. What kind of experience have you had with public opinion concerning the usage of mobile positioning data?

5. What kind of technical procedures/issues have you had when providing mobile positioning data?

E) Are there any advices to the project consortium?

1. Obstacles, hidden pitfalls, challenges, risks:

2. Helpful partners, opportunities:

3. What is your overall opinion about the usage of mobile positioning data concerning all the positive and negative aspects?

Date _____

Place _____

Statistical Offices (National or Regional Level)

This interview is a part of Eurostat's (European Commission) project "Feasibility study on the use of mobile positioning data for tourism statistics" where the key objective is to investigate the possibilities of the use of anonymous aggregated data from mobile network operators in generation of tourism statistics.

The aim of the interview is to understand the aspects of accessibility barriers and opportunities of mobile positioning data in tourism sphere as well as the experiences of organizations like yours.

The project is conducted by a consortium composed of Positium LBS (Estonia), Statistics Finland, Statistics Estonia, French Institute of Science and Technology for Transport, Development and Networks, The Institute for Tourism Research in Northern Europe (Germany) and University of Tartu.

The result of this project will be publicly available by Eurostat from the Summer 2014. If you want to receive more information on the project or on using anonymous mobile positioning data, please let us know directly.

A) Are, to your knowledge, positioning data already being used for the enhancement of official statistics?

1. For statistics in general (if yes: please specify) (STO only):

2. For tourism statistics:

B) How helpful would the enhancement of tourism statistics with mobile positioning data (if feasible) be?

1. Which would be the main fields of application (e.g. forecasting, precision enhancement)?

2. Which would be the main types of visitors to be covered by such a system (e.g. day visitors, overnight visitors, event visitors, international visitors)?

3. What are main requirements for such an enhancement?

4. Which data and data quality would you expect?

C) Are there any advices to the project consortium?

1. Obstacles, hidden pitfalls, challenges, risks:

2. Helpful partners, opportunities:

3. What is your overall opinion about the usage of mobile positioning data concerning all the positive and negative aspects?

Date _____

Place _____

Data Protection Agencies

This interview is a part of Eurostat's (European Commission) project "Feasibility study on the use of mobile positioning data for tourism statistics" where the key objective is to investigate the possibilities of the use of anonymous aggregated data from mobile network operators in generation of tourism statistics.

The aim of the interview is to understand the aspects of accessibility barriers and opportunities of mobile positioning data in tourism sphere as well as the experiences of organizations like yours.

The project is conducted by a consortium composed of Positium LBS (Estonia), Statistics Finland, Statistics Estonia, French Institute of Science and Technology for Transport, Development and Networks, The Institute for Tourism Research in Northern Europe (Germany) and University of Tartu.

The result of this project will be publicly available by Eurostat from the Summer 2014. If you want to receive more information on the project or on using anonymous mobile positioning data, please let us know directly.

A) To your knowledge, are mobile positioning data already being used for one or more of following purposes:

1. … anonymized or non-anonymized identification or tracing of individual mobile devices for public or private purposes?

2. ... applications like traffic monitoring (FCD), police or emergency services etc?

3. …more specifically: e.g. for the enhancement of official statistics?

4. Have you been involved in the process of using mobile positioning data in the aspect of assuring privacy protection of the subscribers?

B) If yes: What do you know about these projects or applications? Please give more detailed information on each of the projects or applications (or name a reference):

1. Which companies/organizations were involved?

2. What data were used? How were they retrieved?

3. Was/Is it a onetime project or a continuous operation? Is it still active?

4. What were the main learnings? Would you work with the same setup again? What would you change?

5. Is there any documentation on the project? Is it accessible?

C) What are, in your opinion, main barriers for the implementation of a positioning service for tourism statistics? Do you see any solutions? Please consider the following aspects.

1. Privacy and regulatory issues:
    a. What are the common procedures (concerning your agency in your country) when mobile network operators want to provide any kind of (aggregated or anonymised raw) mobile positioning data?

    b. Have you had any incidents, problems concerning the usage of mobile positioning data?

    c. How is the privacy of the subscribers guaranteed?

d. What legislation applies in your country concerning the usage of mobile positioning data?

2. Financial and business issues:

3. Technological issues:

D) Are there any advices to the project consortium?

1. Obstacles, hidden pitfalls, challenges, risks:

2. Helpful partners, opportunities:

3. What is your overall opinion about the usage of mobile positioning data concerning all the positive and negative aspects?

Feasibility study on the use of mobile positioning data for tourism statistics
Eurostat contract nr. 30501.2012.001-2012.452

Date _____

Place _____

Research Institutes (Universities)

This interview is a part of Eurostat's (European Commission) project "Feasibility study on the use of mobile positioning data for tourism statistics" where the key objective is to investigate the possibilities of the use of anonymous aggregated data from mobile network operators in generation of tourism statistics.

The aim of the interview is to understand the aspects of accessibility barriers and opportunities of mobile positioning data in tourism sphere as well as the experiences of organizations like yours.

The project is conducted by a consortium composed of Positium LBS (Estonia), Statistics Finland, Statistics Estonia, French Institute of Science and Technology for Transport, Development and Networks, The Institute for Tourism Research in Northern Europe (Germany) and University of Tartu.

The result of this project will be publicly available by Eurostat from the Summer 2014. If you want to receive more information on the project or on using anonymous mobile positioning data, please let us know directly.

A) Are, to your knowledge, positioning data already being used for one or more of following purposes?

1. … anonymized or non-anonymized identification or tracing of individual mobile devices for public or private purposes?

2. ... applications like traffic monitoring (FCD), police or emergency services etc?

13

3. … more specifically: e.g. for the enhancement of official statistics?

B) If yes: What do you know about these projects or applications? Please give more detailed information on each of the projects or applications (or name of a reference):

1. Which companies/organizations were involved?

2. What data were used? How were they retrieved?

3. Was/Is it a onetime project or a continuous operation? Is it still active?

4. What were the main learnings? Would you work with the same setup again? What would you change?

5. Is there any documentation on the project? Is it accessible?

C) What are, in your opinion, main barriers for the implementation of a positioning service for tourism statistics. And do you see any solutions? Please consider the following aspects.

1. Privacy and regulatory issues (specifically for DPAs, but also for other interview partners):

2. Financial and business issues:

3. Technological issues:

D) Are there any advices to the project consortium?

1. Obstacles, hidden pitfalls, challenges, risks:

2. Helpful partners, opportunities:

3. What is your overall opinion about the usage of mobile positioning data concerning all the positive and negative aspects?

Date _____

Place _____

Annex 6. Relevant Patents

Patents for Data Capture and Extraction

Title	Publication number (link) Date Applicants	EU protected	Abstract	Relevance for tourism statistics
Method and apparatus for determining incorrect antenna configuration within a cellular communication network	US2013273921 (A1) 2013-10-17 KENINGTON PETER [GB] RANDELL NICHOLAS JAMES [GB]	NO	A system and method of determining incorrect antenna configuration within a cellular communication network. The method comprises obtaining crowd-sourced data comprising at least one geographical characteristic for a coverage area of at least one cell sector of the cellular communication system, verifying whether the at least one crowd-sourced geographical characteristic for the coverage area of the at least one cell sector is consistent with network configuration data for the at least one cell sector, and identifying a potentially incorrect antenna configuration for the at least one cell sector if the at least one crowd-sourced geographical characteristic for the coverage area of the at least one cell sector is inconsistent with the network configuration data therefor.	The relevance of this patent for tourism statistics is that it allows for a system and method for verifying and correcting whether the raw data obtained from mobile network operator source systems, actually are representative. As such incorrect antenna configurations can be filtered out (and population numbers corrected). This patent adds increased reliability of data.

Analysing audiences at public venues	WO2011080707 (A3); WO2011080707 (A2) 2011-07-07 METERLIVE LTD MUMCUOGLU MICHAEL ENGEL GIORA	NO application withdrawn at EPO	A method for monitoring an audience, includes receiving transmissions over the air, in accordance with a standard communication protocol, from one or more wireless communication devices (24) belonging to members (26) of the audience at a location. The transmissions are analysed in order to derive a characteristic of the audience.	This patent is highly relevant in that basically covers the steps explained in the previous chapter. Especially interesting is that in addition to identifying the wireless devices in a network, these are linked to demographic data. It is relevant to note that the patent was submitted in an application at the EPO, but was withdrawn and as such this method for monitoring audiences as described is not protected.
System and method for using cellular network components to derive traffic information	US2011117896 (A1); US8494496(B2) 2011-05-19 AT & T MOBILITY II LLC [US]	NO	A traffic reporting system and method for geographic area of interest. The system includes standard wireless telecommunication components configured to establish search criteria, determine a sample size, collect traffic information, calculate additional traffic information, and generate reports.	This patent is relevant for the study as it demonstrates the existence of a US patent of determining within a geographic location how much network traffic with how many network devices is generated. In this patent there is not the linkage to demographics or other user data, but movement can be determined over time.
Tempo spatial data extraction from network connected devices	US2011055216 (A1) 2011-03-03 TRENDIT LTD [IL]	NO application withdrawn at EPO	A computer implemented data processing system for estimating an amount of people situated in a specific location and their geo-demographic classification within time range is provided herein. The system is combined of a collector that is configured to collect data on signals and each signal is given a unique ID; an association module configured to associate each signal	This seems to be the most complete patent application ever made in Europe, which is highly applicable for tourism statistics. Here again the application was made at EPO but withdrawn for unknown reasons. Had EPO granted this patent, it would have posed a very restrictive patent to work around –

			with a respective location, namely, place of origin; a processing unit configured to calculate total number of users subscribed to a specific network service provider situated in a specific location and time range; calculate a dynamic ratio by research and statistical data; and an estimation module configured to estimate the amount of people originated from a specific location and the overall amount of people in a location within a time range, by applying the calculated dynamic ratio, that was calculated to each time stamp separately.	since it is that complete.
Method and arrangement relating to mobile telephone communications network	US6587691 (B1) / EP1155590 (B1) 2010-07-14 ERICSSON TELEFON AB L M [SE]	YES (in AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE)	The present invention relates to an arrangement in a mobile communications network, including a first Mobile Station, a second Mobile Station, a Base Station device, control means to control the Base station device, and means for positioning the first and second Mobile Station, the first Mobile Station being a seeking Mobile Station and the second Mobile Station being a sought Mobile Station. The arrangement includes a data storing arrangement for storing position data about the first and second activated mobile stations received from the position means, and means to process the position data with respect to the position of the first Mobile Station and provide the first Mobile Station with location information of the second Mobile Station.	Highly relevant for the first step of the collection of data on the position of a mobile terminal. MNOs that make use of this patent would demonstrate a readiness to take part in a tourism statistics initiative. Of course they would also need to allow linkage of this position data to other user data including demographics and inbound/outbound data.

Standalone positioning in 3G UMTS systems	US2008108374 (A1) 2008-05-08 MOTOROLA INC	NO		
Method of determining position in a cellular communications network	EP1378141 (A1); EP1378141 (B1) 2004-01-07 KONINKL PHILIPS ELECTRONICS NV [NL]	YES (in AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE, TR)	This invention relates to a method of determining position in a cellular communications network which relies on a mobile wireless terminal determining its position on the basis of time of arrival (TOA) measurements on ranging signals transmitted by several base stations whose geographical positions or their antenna positions are known accurately. Such a method is generally referred in the art as cellular positioning	The relevance of this patent is that it shows that there is technology at the disposal of mobile network operators, to optimise the transmissions generated to serve mobile terminals, based on monitoring the position and movement of the terminals. This patent is relevant as it provides a way of understanding how mobile network operators collect data. Those MNOs that use this technology will undoubtedly in some form or format possess data that discloses the geo-movements of terminals. It does however say nothing about user demographics etc. So this could be data that would need to be further linked to other data in order to be useful. But asking MNOs whether they apply this technology can say a lot of the 'readiness' of MNOs to participate in a tourism statistics initiative.

Patents for Data Processing and Analysis

Title	Publication number (link) Date Applicant(s)	EU protected	Abstract	Relevance
Method of system for increasing the reliability and accuracy of location estimation in a hybrid positioning system	EP2635915 (A1) 2013-09-11 SKYHOOK WIRELESS INC [US]	YES (in AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LI, LT, LU, LV, MC,	Methods and systems of hybrid positioning are provided for increasing the reliability and accuracy of location estimation. According to embodiments of the invention, the quality of reported locations from specific sources of location is assessed. Satellite and non-satellite positioning systems provide initial positioning estimates. For each positioning system relevant information is collected and based on the collected information each system is assigned appropriate weight.	Contrary to the technical explanation in the previous chapter this method focuses also on non-mobile network data – hence the term hybrid. As positioning data are generated by multiple devices, this method has a wider scope. Reliance on this method and system as described for generating tourism statistics sees the involvement of multiple data collection stakeholders e.g. MNOs, sports activity monitoring parties, location based service providers etc etc) Using this methods thus increases coordination among stakeholders, but also opens opportunities where MNOs are restricted to provide data, other parties may be less restricted. The cross-linking and verification of multiple hybrid data sources, in theory may improve

		MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR) Examination on-going		the accuracy of data sets.
Method for estimating tourists in scenic spot based on mobile communication technology	CN103020732 (A) 2013-04-03 UNIV SOUTHEAST	NO	The invention provides a method for estimating tourists in scenic spot based on mobile communication technology. The method comprises three parts consisting of a tourist information acquisition module, a data processing module and a decision output module, wherein the tourist information acquisition module acquires various information of each base station of the administrative region in which the mobile users enter, the data processing module performs concentrated processing to the mobile user information to form suggestion and warning instruction, and the decision output module achieves	This method seems very clear and the most ideal for its use in tourism statistics generation as assessed in the feasibility study. However HOW privacy and social ethics are violated is not discussed in-depth. It would be very interesting to understand further how this is dealt with, as privacy and protection of data is a major obstacle in the generation of tourism statistics using mobile positioning data.

			the visual output. By adoption of the method, the tourist distribution information is collected in real time, the provenance and the consumption ability of the tourists are counted without involving individual privacy or violating social ethics, and the reliable decision support is provided for the tourism management department to dredge and guide, tourism recommendation, differential scenic region service and tourist facilities planning and distributing; the count range is not limited at the scenic regions, and participations in any forms of the tourists are not required; and the method is convenient to operate, the manpower and the material resource are saved, and the decision and implementation timeliness are high, thus the method is convenient for developing.	
Method of providing location based service information	WO2013039633 (A1) 2013-03-21 MICROSOFT CORP [US]	YES (in AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES,	An embodiment of the invention provides a location based service (LBS) that updates a first version of information provided a mobile terminal responsive to a geo-query relative to a second, later version of information responsive to the geo-query by transmitting portions of the second version to the mobile terminal rather than all of the second version.	This invention has not just the positioning meta data in scope but a combination of positioning meta data, GPS data, and actual user defined queries. Combining and cross-analysing these then inputs are provided for location based services. So it represents not just an analysis of time-space movements but also user/content inputs

		FI, FR, GB, GR, HR, HU, IE, IS, IT, LI, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR) Application started		and transmission. Its relevance for tourism is that location based tourism services can be provided e.g. local promotions, hotel deals, real-time advertisement.
Method and system for population flow statistics	EP2535843 (A1) 2012-12-19 HANGZHOU	YES (in AT, BE,	A method and a system for providing people flow statistics are disclosed in the invention, wherein multi-types of classifiers connected in parallel are used to	This represents a complete different approach of identifying and counting people based on image-based identification of human heads,

	HIKVISION SOFTWARE CO LTD [CN]	BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LI, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,	perform a human head detection in a current image, the respective human heads in the current image are determined, the respective determined human heads are tracked to form a human head target movement track, and a people flow is counted in the direction of the human head target movement track. A plurality of classifiers connected in parallel are used in the invention so that multi-types of human head targets can be detected simultaneously, such as dark coloured hair, light coloured hair, caps of various colours, and the like.	and distinguishing features including hair colour etc. which could be tied to race-based differentiation. It has limitation in distilling other demographical data. But still race-ethnicity derived data can be produced which provided input for certain segments of tourism markets (...perhaps...it would be a bit controversial too)

		TR) Examination on-going		
Tourism economic data digging, analysing and predicting system	CN202533993 (U) 2012-11-14 UNIV SHANDONG SCIENCE & TECH	NO	The utility model discloses a tourism economic data digging, analysing and predicting system. The tourism economic data digging, analysing and predicting system adopts a C/S architecture and is constructed based on a local network. The tourism economic data digging, analysing and predicting system is composed of a data base server and at least one client terminal. The tourism economic data digging, analysing and predicting system mainly includes five functional modules of tourism income management, per capita expenditure management, sampling data management, analysis and prediction, and statistic diagram output. The tourism economic data digging, analysing and predicting system realises the current state analysis and prospect prediction of tourism economy through establishing mathematic models of statistic data mainly in index classes of tourism industry, thereby providing references for researching and regulating the development trend of the tourism economy.	This is highly relevant for the feasibility study as it represents an end-user / decision maker tool for performing certain statistical modelling and analysis on tourism data. Possibly such a system could be fed by datasets including mobile positioning data, and inbound /outbound data.

A system and method for estimating distribution of outbound roamers	WO2012131553 (A1) 2012-10-04 TURKCELL TEKNOLOJI ARASTIRMA VE GELISTIRME ANONIM SIRKETI [TR] OKUROGLU BAHRI [TR]	YES (in AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LI, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,	The present invention relates to a system (1) and method (100) enabling to operate this system (1) which is intended for estimating distribution and number of outbound roamers of competitor operators and comprises: at least one mobile device (2) which is used the local subscriber; at least one local network (3); at least one VLR (4); at least one SMSC (5) which enables to send/receive SMS in the local network (3); at least one distribution estimation unit (6) which uses messaging traffic that is carried out for SMS's sent by local network (3) subscribers to domestic competitor network (B) subscribers; and at least one database (7) which is a permanent storage media.	While operators can obtain distribution of number of their outbound roamers on operators in any country over their systems, it is not possible for them to learn number and distribution of outbound roamers of other domestic competitor operators. This information can be obtained by requesting it from all operators in a foreign country, however, operators will not provide such information to their competitors. That this method allows for estimating the distribution of outbound roamers, is relevant for tourism statistics, as not always complete data sets are available for statistical analysis, and sometimes assumptions need to be made based on the available method.

		SE, SI, SK, SM, TR)		
System and method for population tracking, counting, and movement estimation using mobile operational data and/or geographic information in mobile network	US2012115505 (A1); US8559976 (B2) 2012-05-10 NTT DOCOMO, INC	NO	Methods and apparatuses are disclosed herein for population tracking, counting and/or movement estimation. In one embodiment, the method comprises receiving mobile phone operational data indicative of user equipment location, where the event data includes location area update messages and periodic registration messages; and performing travel estimation based on the mobile phone operation data, including performing interpolation on data associated with one or more individuals in a population to estimate intermediate positions of a trajectory of each of the one or more individuals for a specified time period based on a shortest path mesh sequence estimation algorithm.	Similar to the traffic information patent, this patent is useful for providing data on quantities of people visiting destinations, and estimating their onwards travel routes, only by looking at the numbers of mobile phone users within a location, and without looking at demographical data. In this sense, to promote tourism, for example access roads can be improved
Method and system for providing traffic and related information	EP1316079 (A1); EP1316079 (B1); EP1316079 (A4) 2003-06-04 CUSTOM TRAFFIC PTY LTD [AU]	YES, granted 09.01.2008 for AT, BE, CH,	The invention provides a system for providing traffic or related information including: a database storing historical traffic data being operable to receive substantially real time traffic data and associated data; means for integrating historical, real time and associated traffic data with respect to traveller profiles to produce customised forecasted traffic information	The relevance for tourism statistics is that destination management information is made available through this. Not necessarily data that allows for demographical segmentation, but certainly access to destinations can be assessed, i.e. where there are structural traffic bottlenecks, etc., which can be improved with

		CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE, TR	with respect to those traveller profiles; and means for sending the customised forecasted traffic information to an intended recipient wherein the customised forecasted traffic information includes predicted travel delays for travel routes described in the traveller profiles.	the aim better managing destinations.

Annex 7. Aggregated Raw Data Example

Aggregation disables the possibility to calculate various important domain estimates, apply algorithms for the identification of individual trips, estimate possible biases like border bias and evaluate processing errors like discrimination between domestic tourism and domestic usual environment. However, aggregated raw data can still provide a significant overview of tourism statistics and are still usable for some objectives.

Aggregation of the raw data can be based on the number of subscribers within space unit in a period of time (e.g. number of subscribers per day per administrative unit). An artificial situation is used to illustrate one possibility to aggregate the mobile data (Figure 16).

Aggregation tables should be created for different dimensions separately from initial raw data and not by aggregating based on each other because such calculation will not always produce logical results (e.g. one subscriber in two administrative territories in one day is presented as 2 unique subscribers on administrative level aggregation but 1 unique subscriber on national level aggregation).

Following figure and table present the examples of different levels of aggregation that can be used for the illustrated artificial situation.

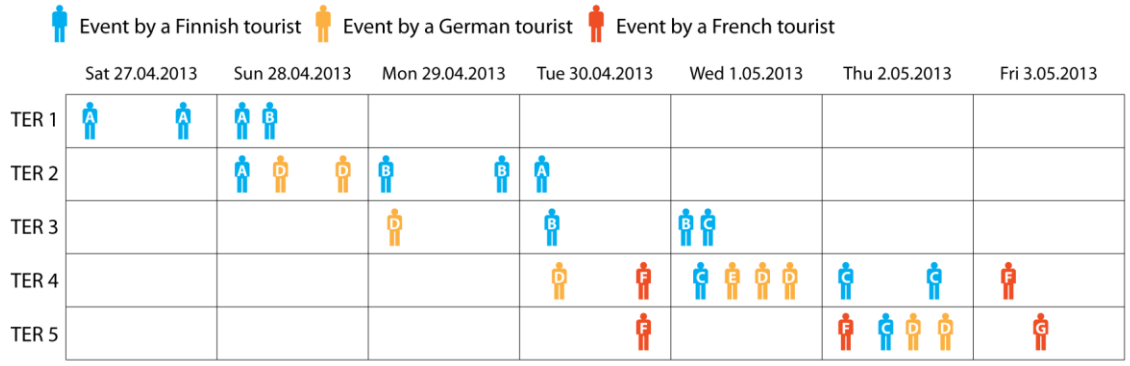

Figure 16. Illustration of artificial situation to describe the aggregation logic. A total of 7 subscribers (A, B, C, D, E, F, G) from 3 countries (FI, DE, FR) registered in an MNO's network in a country's 5 administrative territories (TER 1 – TER 5) during a 7-day period (2 different months – Apr and May 2013).

Table 10. Example of aggregations (presented also in text for explanation).

Aggregation: Monthly, whole nation, no country of origin

	April 2013	**May 2013**
Unique subscribers	4	6
Subscribers per day	10	9

- There were 4 unique subscribers (A, B, D, F) in April 2013 and 6 unique subscribers (B, C, D, E, F, G) in

May 2013 in the whole country;

- There were 10 subscribers per day (4 subscribers were present in 10 days) in April 2013 and 9 subscribers per day (6 subscribers were present in 9 days) in May 2013 in the whole country;

Aggregation: Monthly, whole nation, with countries of origin

Country	April 2013	May 2013
Finland (unique)	2	2
Germany (unique)	1	2
France (unique)	1	2
Finland (subscr. per day)	6	3
Germany (subscr. per day)	3	3
France (subscr. per day)	1	3

- There were 2 unique subscribers from Finland (A, B) in April 2013 and 2 unique subscribers from Finland (B, C) in May 2013 in the whole country;
- There was 1 unique subscriber from Germany (D) in April 2013 and 2 unique subscribers from Germany (D, E) in May 2013 in the whole country;
- There was 1 unique subscriber from France (F) in April 2013 and 2 unique subscribers from France (F, G) in May 2013 in the whole country;
- There were 6 subscribers per day from Finland in April 2013 and 3 subscribers per day from Finland in May 2013 in the whole country;
- There were 3 subscribers per day from Germany in April 2013 and 3 subscribers per day from Germany in May 2013 in the whole country;
- There were 1 subscriber per day from France in April 2013 and 3 subscribers per day from France in May 2013 in the whole country;

Aggregation: Monthly, per territory, with country of origin

	Ter 1		Ter 2		Ter 3		Ter 4		Ter 5	
Country	**April**	**May**	**April**	**May**	**April**	**May**	**April**	**May**	**April**	**May**
FI (unique)	2	0	2	0	1	2	0	1	0	1
DE (unique)	0	0	1	0	1	0	1	2	0	1
FR (unique)	0	0	0	0	0	0	1	1	1	2
FI (subscr. per day)	3	0	3	0	1	2	0	2	0	1
DE (subscr. per day)	0	0	1	0	1	0	1	2	0	1
FR (subscr. per day)	0	0	0	0	0	0	1	1	1	2

Aggregation: Weekly, whole nation, no country of origin

	Week 17	Week 18
Unique subscribers	3	7
Subscribers per day	4	15

- There were 3 unique subscribers (A, B, D) in week 17 of 2013 (starting with Monday 22.04.2013) and 7 unique subscribers (A, B, C, D, E, F, G) in week 18 (starting with Monday 29.04.2013) in the whole country;
- There were 4 subscribers per day in week 17 15 subscribers per day in week 18 in the whole country;

Aggregation: Weekly, whole nation, with countries of origin

Country	Week 17	Week 18
Finland (unique)	2	3
Germany (unique)	1	2
France (unique)	0	2
Finland (subscr. per day)	3	6
Germany (subscr. per day)	1	5
France (subscr. per day)	0	4

- There were 2 unique subscribers from Finland (A, B) in week 17 and 3 unique subscribers from Finland (A, B, C) in week 18 in the whole country;
- There was 1 unique subscriber from Germany (D) in week 17 and 2 unique subscribers from Germany (D, E) in week 18 in the whole country;
- There was 0 unique subscribers from France in week 17 and 2 unique subscribers from France (F, G) in week 18 in the whole country;
- There were 3 subscribers per day from Finland in week 17 and 6 subscribers per day from Finland in week 18 in the whole country;
- There were 1 subscriber per day from Germany in week 17 and 5 subscribers per day from Germany in week 18 in the whole country;
- There were 0 subscribers per day from France in week 17 and 4 subscribers per day from France in week 18 in the whole country;

Aggregation: Daily, whole nation, no country of origin

	27.04	28.04	29.04	30.04	1.05	2.05	3.05
Unique subscribers	1	3	2	4	4	3	2
Subscribers per day	1	3	2	4	4	3	2

- The number of the unique subscribers and the number of the subscribers per day is equal in day aggregation.

Annex 8. Table of Responses by Countries and MNOs

Table 11. An overview of the state of mobile positioning data in European countries, including the situation on data retention, number of MNOs and NSIs reached for this study and usage of mobile positioning data in projects. Note: STAT = Official statistics; RES = Research; BIZ = Business use; PUB = Other public projects.

	Data Retention Directive Source: Arbeitskreis Vorratsdaten-speicherung		**Mobile Network Operators**				**Use of anonymous mobile positioning data** **Source: Survey, interviews, Report 1**			
Country	Transposed	Retention period					NSI interest	Current or past projects, incl. early phases, pilots		
EU-28	Yes / No	For mobile data, months	#	Contacted	Replied	Discussed	For official statistics	Tourism	Mobility and Transport	Other
Austria	y	6	3	3	1	1	Y	STAT	BIZ, RES	PUB
Belgium	N	N/A	3	3	2	1	N/A	RES	BIZ, RES	RES
Bulgaria	Y	12	5	5	0	0	N			PUB
Croatia	N/A	N/A	3	3	0	0	N/A			
Cyprus	Y	6	2	2	0	0	N/A		RES	
Czech Republic	N	N/A	5	5	0	0	Y	BIZ	BIZ, RES	BIZ
Denmark	Y	12	4	4	0	0	Y	STAT (Greenland)	RES	
Estonia	Y	12	3	3	3	3	Y	STAT, BIZ, RES	BIZ, RES	BIZ, PUB, RES
Finland	Y	12	4	4	3	3	Y	RES	STAT	
France	Y	12	4	4	2	2	N/A	RES	RES	RES
Germany	N	N/A	4	4	3	3	N		BIZ, RES	BIZ, PUB
Greece	Y	12	3	3	0	0	N/A		RES	
Hungary	Y	12	3	3	0	0	Y		RES	
Ireland	Y	25	5	5	0	0	Y	STAT	RES	RES
Italy	Y	24	4	4	0	0	Y		BIZ, RES	RES
Latvia	Y	18	4	4	1	0	Y		RES	
Lithuania	Y	6	5	5	0	0	N		RES	
Luxemburg	Y	6	4	4	0	0	N			
Malta	Y	12	3	3	0	0	N			
Netherlands	Y	12	3	3	3	3	Y	STAT, RES	STAT, BIZ, RES	RES
Poland	Y	24	4	4	1	0	N/A	BIZ, RES	BIZ	
Portugal	Y	12	3	3	0	0	Y		RES	RES

Romania	Y	6	5	5	0	0	N/A			
Slovakia	Y	12	3	3	0	0	N/A			
Slovenia	Y	14	4	4	2	2	Y	STAT	STAT	
Spain	Y	12	4	4	0	0	Y	BIZ, RES	RES	BIZ
Sweden	Y	6	5	5	0	0	Y	STAT	RES	RES
United Kingdom	Y	12	4	4	1	1	N/A	BIZ	RES	BIZ
EFTA										
Liechtenstein	Y	6	3	0	0	0	N/A			
Norway	Y	6	4	0	0	0	Y			
Switzerland	Y	6	3	0	0	0	N/A		BIZ, RES	RES
Candidate countries										
FYRO Macedonia	N/A	N/A	3	0	0	0	N/A			
Iceland	Y	N/A	4	0	0	0	Y			
Turkey	Y	6-24	3	0	0	0	N/A			
Montenegro	N/A	N/A	3	0	0	0	Y	STAT		
Serbia	N/A	N/A	3	0	0	0	N/A			

Annex 9. Analysis of the Survey Results and Interviews with Stakeholders

In the survey as well as in the interviews the main topics emphasised are the regulatory frames and limits that MNOs can act within (in providing access to the data), the essence of the data that can be used in tourism statistics (the format of the data, level of anonymity), technological processes that MNOs require to be implemented in order to provide such data, financial limitations, business secrets and MNOs' internal interests for the use of this data.

The results of the online survey and expert interviews are combined and analysed together as they complement each other. The questionnaire alone would have been too thin to provide a comprehensive analysis as there were many questions structured to be multiple-choice. The possibility to expand with comments were not often used. The expert interviews, on the other hand, went more in-depth in some of the key subjects.

Data Availability and Accessibility

The Potential and Use of Mobile Positioning Data

According to the survey results 89% of the respondents are using tourism statistics, which indicates that the survey is delivered to appropriate experts and stakeholders. 86% of the respondents are aware of the possibilities of using mobile positioning data. But despite the high awareness of the possibilities provided by mobile positioning data, mere 14% actually use it.

Mobile positioning data is used for the most part by research institutes (8 respondents) and to a lesser degree by private enterprises (see Figure 17). The survey discovered very few statistical offices and municipalities using such data. Even if the organisations are interested in the usage of mobile positioning data, they haven't got any access to it yet.

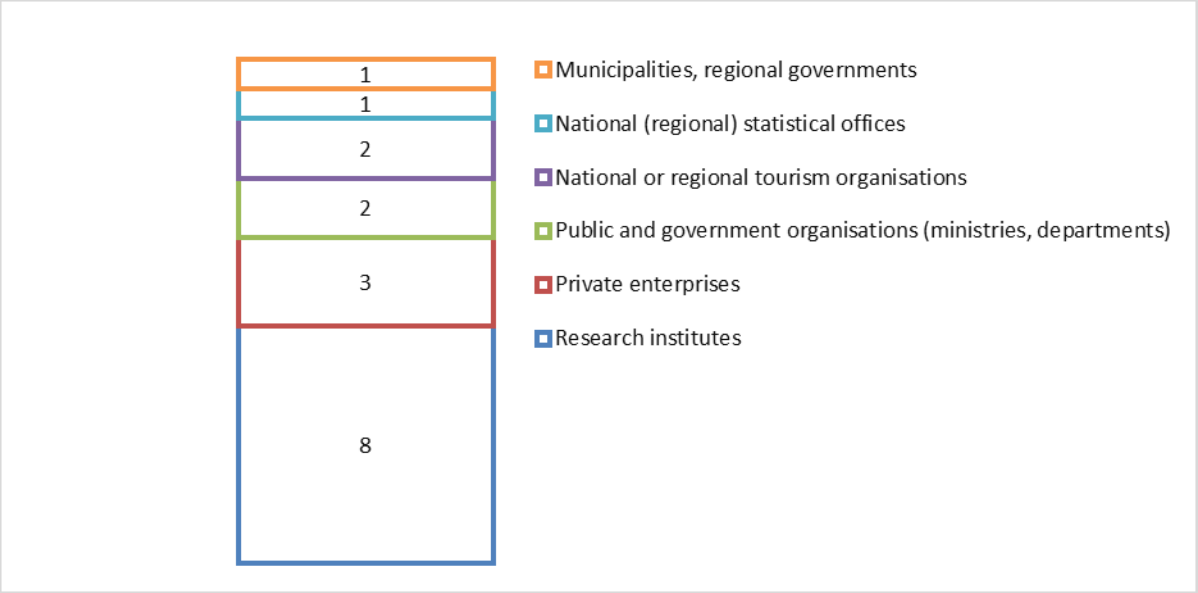

Figure 17. Survey results on type of organisation that use mobile positioning data (n=116).

As the survey shows, mobile positioning data is used in Spain, Netherlands, Czech Republic and Italy. In the Netherlands, initial steps are being made towards the use of the mobile positioning data to compile national statistics from aggregate data while respecting the privacy of users. In Hungary, a PDA tool is used in the tourism statistics section in case of two statistical activities: the expenditure of non-resident visitors and resident visitors abroad. First steps have also been taken in many countries to start using mobile positioning data in tourism statistics, e.g. in Ireland, Slovenia, Belgium, Austria and Greenland. In Ireland, the usage of mobile positioning data is very possible in the near future. An overview is provided in Annex 7 as to which countries are starting or have had projects, and which NSIs are interested in using mobile positioning data in official statistics in the future.

There is an on-going evaluation process also in Slovenia about the possible use of this type of data and designing the draft processes. A project is planned in Belgium to monitor visitors to open events, negotiations with a mobile operator is on-going. In Greenland, the mobile positioning data usage has been discussed as a means of locating different nationalities in Greenland, but the legal implications are not yet fully analysed. In Austria, the first steps have been taken by exploring technical and legislative possibilities with one of the main mobile providers in the country. Previously, in 2011, the statistical office wanted to get mobile positioning data for the production of tourism statistics, but unfortunately the Austrian MNO pulled out of this project and at present is no longer offering mobile positioning data. Probable reasons for this were public opinion/data protection issues raised by the press. In

France, there are several researchers that have used mobile phone data and they have used volunteers to study their mobility using a combination of mobile phone data and survey data.

Mobile positioning data is usually received either directly from mobile network operator(s) (29%) or through other channels (35%) such as obtaining the data directly from phone-owners (GPS and specific geolocation applications). In fewer cases the data is received from the research institutes (18%) and from mobile positioning data providers (other than MNOs) (18%). It is a fundamental assertion and confirmed also by the online survey that for obtaining the data a lot of personal connections need to be put into effect, which makes the communication and interaction between different parties and organisations essential for a sustainable operation. The procedures needed to be followed to obtain the data have similarities in all the responses in the survey: (IP) agreements, contracts or experiment agreements had to be signed between the MNOs and the organisations using the data; the licenses and specific software is purchased and sometimes provided to users; in a few cases the organisations using mobile positioning data have their own software installed in their mobile phone panels. There are cases when the organisation is being contacted themselves by mobile data provider, who introduces these data for tourism monitoring, which has led to the signing of the contract on the basis of the tender.

Usually only one MNO from the respective country has provided the data, but it was also common that data comes from several different sources (apart from directly from MNOs), including direct data from mobile devices (small sample, special software, avoiding MNOs). It was also specified in several responses that the negotiations with MNOs are in progress. Use of mobile positioning data has a high level of secrecy, which is why 35% of the respondents were not allowed to name the MNOs and some of the users of data do not know the origin of the data.

The data expected from MNOs are commonly the monthly indicators on the number of subscribers, classified by nationality, by length of stay and by means of transport. From a statistical quality point of view, the main expectation is to have eliminated duplicates. Also vital is the representativeness of data (i.e. not a sample but census), timeliness (faster data compared to traditional sources, possibility of near real-time), and classification of subscribers (country of residence, transits, duration of stay etc.). Data that has been used is related to mobile devices identified on the zones of interest, according to the nationality of the prefix number, in order either to split individuals between local residents, excursionists, tourists and travellers in transit, or to analyse the duration of stay, or to estimate the impact of

specific events. In short, mobile positioning data can be characterised by different aspects like its time-period, amount, content and format. As per the survey there are different formats in which the data arrived, some of which are:

- One week of event data (call, SMS, data transfer) for subscribers;
- Excel (pivot tables) time period (e.g. one year);
- Daily statistics (inbound by country of origin and domestic);
- Weekly and monthly, hourly basis, data of about 3 million users;
- Excel tables with arrivals per country over the summer season;
- Definition of profiles: residents, tourists, excursionists, transit;
- Data indicating coordinates and time (every 20 seconds).

The respondents' organisations have also used inbound roaming and other types of data (e.g. GPS precision directly from mobile device applications). A respondent also mentioned data collection through geo-referencing demographic surveys during field work through interviewer's terminals. It is common to use external parties to collect the data. In half of the cases MNOs provided limited pseudonymous raw data. Aggregated data is also common (25%). Raw data without any anonymisation or aggregation is being used in 6% of the cases. In terms of continuity, the most common case is that the data has been provided through one-time projects and experiments (70%) and therefore there is a lack of continuous data flows. Only occasionally the data has been reproduced a second time with a renewal of the zones and events of interest.

Mobile positioning data is mostly being used for research purpose, but it is also being used to assess the possibility to compile official statistics and for business purposes. The most common field for the use of mobile positioning data is for generating tourism statistics (88%) (see Figure 18). Although this is certainly also caused by the respondents' profiles (mostly tourism-related organisations were involved in the survey). Data is also used in urban planning (47%), in transportation planning (29%), traffic monitoring (29%) and migration research (29%), in commuter sciences (24%), in crisis management (12%) and in geomarketing (6%). Some organisations use mobile positioning data in advertising, geo-referencing demographic survey observations and creating mobility patterns in destination. In Figure 18 it is shown both, fields of data usage by users of mobile positioning data and fields of data usage by those who are interested in mobile positioning data, but in some reasons haven't got access to it.

Those survey respondents who are not currently using mobile positioning data were asked about their interest in using this type of data in their work in the future. A relatively high percentage of respondents (64%) who are interested provides great opportunity to widen the range of potential users and make this type of data more widely available. Only 9% of the respondents not using mobile positioning data expressed that mobile positioning data has no value to their organisation. Although the interest in using mobile positioning data is high, only 17% of the survey respondents have approached mobile operators or other suitable organisations for the possibility of obtaining it. Main data providers are MNOs, however there are other important parties involved as well: private companies, large telecommunication and IT companies, telecommunication agencies, information commissioners, a company providing Bluetooth data, consultancies and app developers.

The main interest for those organisations who are the potential users of mobile positioning data would use it for generating tourism statistics (89%), for geomarketing (26%) and for traffic monitoring (26%). They might also apply this data in transportation planning (21%), urban planning, migration research and commuter studies are the fields less important (all 5%). In Slovenia for example tourism statistics are integrated in general statistical process. There were also mentioned travel statistics and geo-referencing demographic surveys of where to use the mobile positioning data if it were available.

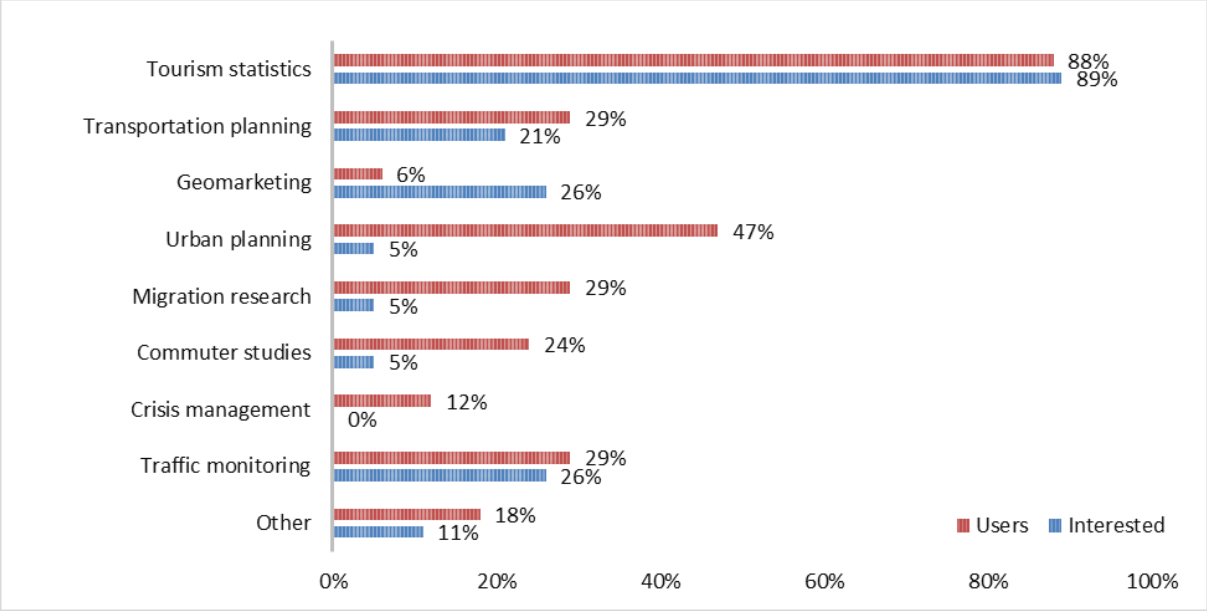

Figure 18. Fields of mobile positioning data usage according to the responses from the survey (users n=17; interested n=75).

It was pointed out that the main issues for not using the data are financial. Often it is not possible to prolong the one-time projects, because of the budget limitations.

Comment from the interview: *'The aim of the pilot project was to try new possibilities of measuring tourism flows and to present it to the destination managers and other stakeholders. As the destination managers can see the effect of such monitoring, some of them decided to continue in monitoring on their own costs.'*

There is a lack of knowledge on the reasons MNOs do not provide data. It was said that the data has been received from an intermediary so that there is no connection with the MNOs. In Spain for example, it is said that there are unprecedented uses of geo-referenced databases and that the events databases are of limited efficacy for statistical purposes in tourism.

Interviews with MNOs indicate that, if the interest in usage and need for mobile positioning data is high, MNOs are reasonably interested in discussing the possibilities to provide the data. But there are several obstacles that MNOs and the users of the data must take into account.

MNOs are mostly sympathetic to the idea of using mobile data in tourism statistics – this topic and the value from this methodology are well understood. Mobile MNOs themselves use mobile positioning data for their internal analyses for fraud detection and customers profiling, also as an input for roaming department and business analysts. In Germany for example, one of the MNO's develops and markets self-learning systems of analysis that use Big Data sources to structure and pattern connections to the benefit of companies and organisations. The attractive combination of a start-up that is powered by investment guarantees an exciting start into the dynamic Big Data market. Another example of the usage of mobile positioning data by MNOs comes also from Germany (Telefonica). They are working intensively on Big Data solutions and in this respect also on the usage of data on location. But in the end they are only able to prioritise subjects when they have a good effect on the balance sheet. For that reason it would be interesting for them to know how cooperation with Eurostat, NSOs etc. would look like in detail.

One of the purposes of MNOs providing the data to other stakeholders for external use is to provide data for governments based on request for official use and also to license the data to brokerage partners. According to MNOs the data is typically used to conduct analyses of roaming trends for explorations; evaluating the possibilities for creating new revenue streams, for example. Mainly, however, mobile positioning data is used in generating official government statistics and for research purpose. The data is being used in generating tourism statistics and also in the fields of transportation planning, geomarketing, urban planning, crisis

management and traffic monitoring, also for creating roaming trends and general travelling trends. In France, for example, business projects for providing daily flows and tourism indicators as for developing countries project on urban planning and transportation, are in the final phase. It is also used for getting an overview of the use of services and for marketing purposes (also included customer base marketing). Not less important is the purpose of analysing market share and for describing network performance. So it can be said that there are strong business interests for providing any kind of services based on the data from MNOs. While financial gain was often named as a motivation for providing the data, getting knowledge such as market share information or other visibility can also be considered highly important for MNOs, to make them willing to give out the data.

One of the MNOs in France for example provides the data in the framework of research agreements and in that case they are not data sellers (as for example Airsage in the US), but already for several years they have been collaborating in several bilateral or multilateral research projects using CDR or signalisation data as well as in collaborative projects. So the interest and benefit of the use of the mobile positioning data can also be considered to be multilateral. It was said that it seems to be rather a hardware/technological cooperation – the possibilities for data analysis are (if anything) only a secondary by-product.

MNOs are mainly asked to provide anonymous billing CDR, but opt-in mobile phone data is also used, along with data from technical probes (network signalisation) in France being mentioned once. Data in raw format without any anonymisation or aggregation is usually provided for a very short period. But there are also some MNOs who offer exactly the opposite: fully anonymous and aggregated CDRs. Signalisation data or CDR is a very simple structure: alias (pseudonymous identity), timestamp, *cell_id* (geographical reference), type of event or sometimes the duration of the connection or volume.

According to MNOs all the information associated with mobile positioning data can be considered a business secret in nature and there are clear internal privacy/compliance rules and policies to be followed when using or providing mobile positioning data. But all MNOs evaluate that there is huge potential for this field in the future since ‘applications are endless.’ However, the MNO’s position here is to provide pre-treated/pre-processed information as possible (data value-add) for specific issues (e.g. indicators for tourism, OD matrixes for population flows), instead of providing the raw data itself.

Although the MNO does not provide individual localisation data to every asker in fear of legal ramifications from the personal data protection act, aggregated datasets can usually be

offered after the approval of a competent national agency. In France, at each stage in providing the data, permission from data protection agencies has to be acquired – moreover, at this moment as the European personal legislation is being transformed and the old rules have to be adapted for, the MNOs have to discuss every new project in detail with the data protection regulator. It was mentioned that (in Germany) every single project has to be approved by the authorities so there is a constant discussion going on. Because of the strict data protection regulations and the sensitive public, it is a big challenge for MNOs to provide services based on location data. The accessibility and availability of the mobile positioning data is then understandably difficult concerning.

The responses from DPAs interviewed concerning their knowledge of the use of mobile data differed in different countries. Usually DPAs just do not know about the cases as they are either not involved or the projects have not been continuous.

Finnish DPA has been involved in previous projects with the Finnish Road Administration (currently Finnish Transport Agency) and former Radiolinja (currently Elisa) where short-term mobile data was used in transportation research. Swedish DPA has been informed of several one-time projects with anonymous mobile positioning data. All in all, it can be said that DPAs interviewed are not aware of the usage of mobile positioning data and have not dealt with the related big problems thereof.

Perceived Value from Mobile Positioning Data

Respondents of the survey and interviews shared their opinion about what problems are solved and what value it brings when using mobile positioning data. The most common perception is that the current data sources for statistics will be considerably enhanced regarding the temporal and spatial accuracy and the possibility to generate new statistical indicators will be examined. The mobile positioning data would be used for providing statistical indicators for visitors of open events such as count, demographics, resident / non-resident tourist, approach route, and visitor management. The main types of visitors to be covered by such a system as the mobile positioning data would be the same-day and overnight visitors and also the volumes of transit travellers, event visitors and business travellers. Such data is reliable for overnight-stays, big events, local mobility analysis, but further information such as expenditure and other behaviour would still have to be collected by surveys.

Comment from the interview: *'If it is considered that mobile positioning data could be used to estimate with precision the total number of visitors /.../, in practice such a system*

would mainly help us to improve our counting of visitors crossing borders by road. Indeed we consider that we have a satisfactory statistical precision in the counting of visitors in others means of transport, and our main topic of improvement is concentrated on the estimation of the number of visitor entering the country by road.'

Comment from the interview: *'Mobile positioning data can also be used as a consultancy equipment and data generator. The current problem statement refers to port of entry, places visited and proxy for country of residence.'*

Several survey respondents who are interested in using mobile positioning data, but have not received it yet, have pointed out that mobile positioning data would be a valuable tool for plausibility checks of tourism data, mainly tourism flows with regard to same-day visits. Also, mobile positioning data would be used as a complementary tool in regard to regular tourism surveys and estimates. It can likewise be used for measuring tourism 'hotspots', looking at how different destinations are co-consumed. The value of mobile positioning data also comes from the dynamic nature of updated traffic information, improved quality of statistics, faster data generation, and reduced respondent burden. Also, mobile positioning data would reduce the burden for the reporting units, at the same time making more data (big data) available to be used for new technologies and processes. In similar fashion, there is quicker and more access to detailed data on the duration, destination and distribution of the stages of inbound tourism trips per country of residence.

Respondents mentioned that MNOs have a large database with information rich for statistical purposes because of the collection of passive mobile positioning data. Important data can be gained about a visitor's travels, geographical information and number of days or nights spent. Using the data gives the opportunity to be able to identify the character of visits on small scale areas important for tourism like UNESCO sites, mountain areas, spa towns, etc. The data is also necessary for scaling unprompted travels, duration, destination and distribution of inbound tourism trip stages per country of residence, understanding tourism behaviour in touristic destinations aimed at the improvement of the destination management system. It helps the data users to understand tourist mobility and the length of the questionnaires will be shorter when mobile positioning data is available. Though this database does not give any information about the visitor's motivation (leisure, work or conference) or travel costs, mobile positioning data would help with providing lower bounds on the number of trips, rough speed estimates, and tourist profiles. It also gives extra statistical information to possibly help correct other data from survey sources, for instance on the average lengths of

stays per origin or area, for seasonality analysis, and it also provides a new tool for the measurement of excursionists in year two and beyond – a quick (instant) information source to assess the follow-up to a period or event.

Mobile positioning data also gives the opportunity to test the new spatial network algorithms. One respondent expressed that for some kind of research mobile data is basically the only possible source. Lower cost of the data collection and analysis cannot be underestimated. A respondent from Belgium also said that the value in their organisation comes from working together with and IT consulting and services company and being able to ultimately link the mobile positioning data to credit and debit card expenditures. In Greenland for example, the usage of mobile positioning data would most importantly reveal tourism mobility patterns within Greenland segmented by nationalities. Secondly, it would present valuable information on the use of mobile phones in different places around Greenland and thus form the basis for disseminating information to mobile users according to their location.

The main requirements for such an enhancement would be the compatibility of 'old' and 'new' methods and definitions. Mobile positioning data has to fit into the established toolset of compiling tourism statistics, such as border surveys, interviews, and accommodation statistics. It is also necessary to be able to determine if a visitor enters the country by road or by another means of transport. Recent exchanges in France suggest that this is possible by using trajectory analysis. The enhancement of tourism statistics with mobile positioning data (if feasible) would be quite helpful in tourism/mobility statistics according to statistical offices, but probably not as a replacement but as a supplement to the existing data (in e.g. plausibility checks; short-term trends; tourism; traffic/mobility). It could also be used for the travel item in the Balance of Payments: (short-term) data on the volumes of daytrips, border crossings. The use of mobile positioning data would improve the estimation of the number of visitors to the country and thus the precision of the border survey. Taking all of these opinions into account, the opportunity to use mobile positioning data has a real value to different organisations.

One aspect that has been emphasised by the respondents is the timeliness of the data i.e. the opportunity to provide fast statistics up to almost real-time. This would increase the confidence of the users of the tourism data even if the fast data is just indicative in the beginning and will be revised later (for more precise indicators).

Although mobile data can be considered only one of many data sources, mobile positioning data cannot be underestimated. The knowledge of such data is growing all over

the world and MNOs are more and more interested in the value they provide by sharing the data they possess. The barriers they need to cross will be examined in the subsequent chapters.

Barriers

Even if the value of mobile positioning data is much appreciated, there are still some barriers for why the data is mostly not given out. Some of the respondents mentioned that they are in the process of acquiring the data. There was also a case where the operator ultimately stopped the process because '*they didn't want to be mentioned or related to the use of their phone data for research other than improving phone facilities*.' – comment from the interview. Sometimes, the data was incomplete or access was granted to only a certain kind of data: aggregated number of SIM cards by month, for example. Several respondents mentioned that the sample or test data was provided and received, but there are still some further activities that need to be discussed. High cost of the data is also a problem limiting access.

The main barriers to obtaining the data from MNOs are privacy concerns and legislation (see Figure 3). The respondents also have to take account of possibly exposing business secrets, high implementation and maintenance costs and finding solutions to different technological problems.

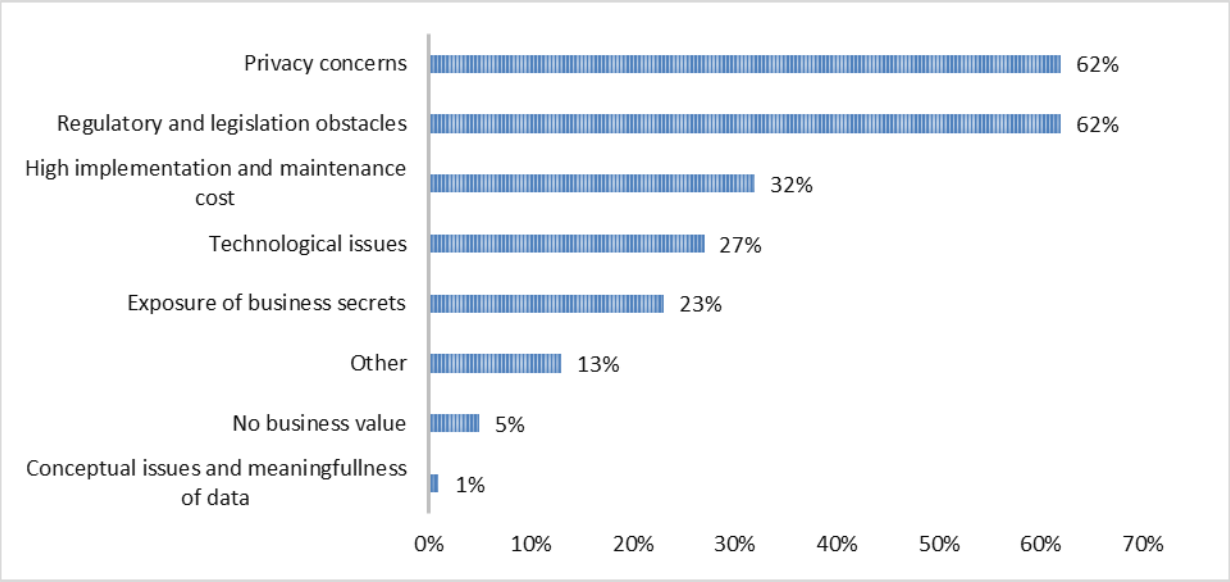

Figure 19. Main obstacles in getting data from MNOs according to responses from the survey (n=116).

It is also clear that getting access to mobile positioning data requires completing specific procedures and dealing with regulatory issues. Among the main regulatory obstacles mentioned were that the new legislation is not yet enacted, that there are some methodological

and security issues, administrative procedures, also legislation governing statistics and data protection, financial constraints as well as institutional problems that must be taken into consideration.

Although the usefulness to explore the feasibility of the use of mobile positioning data for tourism statistics purposes is understandable, several methodological issues are envisaged. Respondents mention issues in relation to the compilation of statistics, such as the increasing number of travellers who own and use more than one cell phone. It is difficult to assess the errors resulting from personal patterns of mobile phone use (people using several numbers and some population segments, like the children and elderly, not using a cell phone at all), so respondents assume only conclusions concerning mobile phone traffic can be made. Also the mobile data does not provide information on expenditure. There are more issues, i.e. border problems, roaming issues, privacy concerns, negative reactions in the media, lack of common platform across operators, general weakness of methodologies used and research questions posed, the high price of data, no qualitative information on the users. Risks concerning the new European data protection regulation and the consistency mechanism were mentioned. Effects on the public opinion were also mentioned as one of the major challenges in these kinds of endeavours.

Storage capacities of the data, statistical compilation, organisation public image, clashes of interest (marketing opportunities of MNOs), concept of usual residence and usual environment, no experience with big data processing, cost of the implementation, coverage issues, and share of the market for each MNO were also among the concerns from the respondents. Drawbacks like people's knowledge of IT would also be among the main barriers in some countries. Also, loss of consumer trust in mobile services due to the intrusiveness of the tracking being undertaken can be a major issue. When data is initially collected, they are far from anonymous, but the higher the level of data aggregation, the less useful it is.

One of the major problems mentioned is the underreporting for nationalities where old GSM technology is used, which is not compatible with European MNOs (mainly American and Asian nations), and the uncertainty of estimation from MNO's market share to the total in inbound roaming market.

It was also said that the diversity of actors is an obstacle. Misleading communication about existing technological and juridical solutions is increasing the complexity of getting adequate technical and financial support for novel projects toward smart solutions.

Privacy and Regulatory Issues

Privacy and regulations are the largest problems reported by the respondents and mentioned during the interviews. This along with the fear from the negative publicity is the major obstacle to overcome.

Comment from the survey: *'It can be scary to imagine a knowing-everything-government, especially about foreigners due to the political context. How could we ensure against abuses while tracking foreigner's mobility and not only tourists?'*

The solution provided is to work on an aggregate scale or to have robust methods allowing the anonymisation of the data. Another aspect is the principle of minimisation. The questions that have been raised are: is it less sufficient? Does the statistics have to be based on exhaustive data or is a sample sufficient? No generic answer can be given for what is processing of truly anonymous positioning data. If pseudonymous positioning data is combined with other linkable surveillance data, the combination immediately becomes personal data. As far as location data is concerned, the possibility to indirectly identify the subscriber depends on the size of the geographical grid used. The data can be personal if the grid is too small. Each operator knows the number of roaming clients in a certain given timeframe. For example, the operators could certainly report statistics on the number of roaming subscribers separately for each nationality, along with duration of visit based on the roaming time. Such statistics would not be considered personal data and the operator is certainly capable of providing this. This response (duration of visit) however is not correct according to the Data Protection Directive's provision disallowing any processing of the data if it is not fully anonymous.

There was also a general question raised: what is meant by data protection overall? According to the recital of the Data Protection Directive, data protection systems are meant to serve people and not the opposite. When systems are developed, the fundamental rights of people are to be taken into account, especially the right for privacy. Fundamentally, data protection means that every person has the right to decide whether he or she wants to be included in Statistics or not and whether they keep their phones on or not. As well as every person have the right to decide whether they have a prepaid, confidential or public subscription etc.

Public Opinion

Because of privacy issues there is a risk for hostile public reaction concerning the use of mobile positioning data. This is why one needs to explain carefully how the actual use of this data is anonymous and secure and what kind of social benefits it could provide. MNOs have sometimes experienced many fears in the case of 'paranoiacs' and low consciousness in the case of 'optimists'. Usually the press and some of the PR-seeking persons in the DPAs can get quite touchy when data protection issues are even only slightly affected. The sensitive public and press can have a big impact. That is why it is especially necessary to show that thanks to big data additional value can be generated for the public.

Financial and Business-related Barriers

One of the major issues of on-going discussions with mobile phone operators is the cost of data. Who is paying for the data, for work-hours, for implementation? What is the value for the final user? These questions are not easily answered and have not a clear response yet.

Main business issues are that the data is owned by MNOs and that they don't 'give' it freely to researchers or to state agencies. A new avenue to monetise data is very attractive for an operator (because of opportunity to sell network data produced and stored anyway). In the end, some of the MNOs are only able to prioritise subjects when they have a good effect on the balance sheet and do not harm the main business. However, the Eurostat study is seen as a possibility to set a good example with the help of the public impact. This might smooth the way for later and stronger commercial projects. Some MNOs already provide data for commercial usage (Telefonica, Orange, etc.) but only in very specific domains and with just a portion of the value from the initial data because of the legislation limitations on processing the personal data (aggregated data, not pre-processed).

Technological Barriers

When providing mobile positioning data some of the technical procedures/issues the MNOs have had is the concern that with sizeable European countries the size of databases can be very large. Good storage and computing technologies are also needed. Also, the respondents suggest that many different specialties are required within the chain of processes that are rare or very costly. The question of internal development versus outsourcing is also mentioned.

Technically, it should be possible to measure the volume, the direction of movement and the nationality of mobile phone users in a given area at a given time. But what needs to be considered is what, when and how often is one measuring it. Will it be done with CDR or other sources of data, what about border bias, etc.?

Possible source of information mentioned concerning the intended use of mobile positioning data is the home and visitor location registries (HLT, VLR – actual antenna of every mobile phone that is connected to the network at a given time). However in this case, no storage of historic data is usually allowed. CDR from billing database (stored for 6 up to 24 months).

If data is anonymised prior to further processing, it should take place continuously and instantly after the appearance of the data in the network. The created anonymous data has to be stored at a different place than the original data. A viable way of anonymisation is the instant erasing of all elements of the data that could be used to identify the user. Only then the data would be truly anonymised with no pseudonymisation e.g. via a random number (token) substituting clear ID elements, leaving only nationality (country code), time and location. As this is the most common anonymisation option, it has to be done in case there is obligatory requirement to anonymise the data prior to processing. However as this limits significantly the value of the statistical indicators, the lack of a good anonymisation algorithms that preserve the valuable properties of the data, can be considered as technical limitation.

In conclusion, there were quite a few technological issues mentioned in the interviews, concerning mobile phone data usage. These include the big volume of data and its storage, also filtering and processing of data, finding the appropriate algorithms, visualisation of tourist flows, etc. Nonetheless, it was said that the technology is relatively mature and is lesser of a problem compared to the legal obstacles.

Legal Obstacles

Responses from the survey and interviews indicate the divergence in the regulations in different countries. The EU umbrella directives allow Member States to implement the legislation based on local peculiarities and thus they cannot be treated the same. In different countries the limitations and the concepts are different both on the privacy protection standards as well as the interpretation of the national statistics acts. Even countries with similar legislation interpret the use of the data differently – in Germany, any kind of mobile

data usage has to be approved by the data protection agency, in Austria the data can be used without approval in some cases.

The power of the national statistics acts is also different. It was mentioned that in some countries the NSOs can already now requisition of the mobile data without amendment in the legislation. In most countries though, the NSOs can requisition only the data that is specified in the statistical act – meaning NSOs usually can and will not act unless there is obligation to requisition the data from MNOs. However the form of the data that can be demanded is almost never clear (for the specifics of MNOs) – whether it is raw personal data or pre-processed and aggregated data.

It is however clear that MNOs are not allowed to process personal data for the purposes other than those mentioned in the contract with the subscribers, for direct business purposes (billing) or based on a request by the state according to local legislation, unless the data is anonymous. It is not clear whether MNOs would have to provide raw personal data or pseudonymous or anonymous data; if they have to pre-process the data themselves (using again raw or pseudonymous or anonymous) or provide un-processed data (prepared data).

Also, the definition of anonymity varies greatly. In some cases, simple decoding the data (pseudonymisation, tokenisation) is considered sufficient, however in some cases very specific criteria are mentioned and the concept of 'everything that is not explicitly allowed is forbidden' prevails.

The experts, including the DPAs, could specify on the legal status of the use of mobile positioning data in four countries – Germany, Austria, Denmark and Finland.

In Germany and Austria the legal basis for mobile positioning data usage is based on the following: EU Data Protection Directive of 1995 (DPD), the concurrent Article 29 Working Party discussions; the telecommunication laws; the Data Protection Acts. One of the big issues concerning the legal basis for the usage of mobile positioning data is the principles of free EU markets vs. needs for data protection.

The Austrian Data Protection Law of 2000 is currently translating the DPD into national law where §46 regulates the privileges of scientific and statistical use of sensitive data.

In Germany, the special directives in the Telecommunication Law state that mobile positioning data falls under the secrecy of telecommunication. Federal authorities have the responsibility to regulate telecommunication.

There are also plans for a new European Regulation, updating the DPD – timing quite open at present. One of the goals is to reach a common understanding of legal terms that up to now are lacking a precise definition. Austria suggests implementing the concepts of 'indirect personal data' and privileges for scientific and statistical use of sensitive data in the new DPD.

Several general rules can be distinguished for Austria. Firstly, according to the Austrian Data Protection Law, MNOs are free to use their data without any concerns from the data protection side, as long as the initial personal data is anonymised. No need for approval by regulation authorities for projects that use data in this sense. Secondly, with regards to appropriation, MNOs are only allowed to use 'personal data' to conduct the services that are agreed on in the contract with the user (i.e. provide telecommunication services and billing for these services). However, there are some exceptions for emergency services, law enforcement (after judicial approval) and police/security (so-called 'Gefahr im Verzug'). Thirdly, aggregation in this context is a process of anonymisation. Austrian law provides a special definition for 'indirect personal data'. This is pseudonymised personal data that cannot be traced back to the original person without breaking the law. That means that there should be no legal/regulatory limitations to using mobile positioning data for tourism statistics nor should the intended use be subject to approval by the regulatory authorities. Fourth, the use of personal data in scientific/statistical projects is subject to approval by the regulatory authorities. Such approval is dependent on the following: final results do not contain personal data; persons have agreed to the use of their personal data (opt-in); the results are in public interest; there is scientific expertise. As a fifth and final point for Austria, approval by the regulation authorities does not oblige the MNOs to cooperate.

Much the same way, some general rules can be distinguished from interviews in Germany. For Germany, everything that is not explicitly allowed is forbidden. Secondly, with regard to appropriation, MNOs are only allowed to use personal data to conduct the services that are agreed on in the contract with the user to conduct the services that are agreed on in the contract with the user. In order to use mobile positioning data outside of appropriation, the data needs to be anonymised. Pseudonymisation is not sufficient enough. 'Anonymisation' by the definition of the German regulation authority is reached when the effort to trace back the anonymised data to the original person is immoderate and economically not viable. Technically, it has to take place continuously and instantly. The new anonymised data has to be stored at a different place than the original data. For the DPA, a viable way of

anonymisation is the instant erasing (not substituting) of all elements of the data that could be used to identify the user.

In Denmark, the two relevant laws concerning the usage of mobile positioning include the Personal Data Protection Act and the Data Retention Executive Order. The first determines the conditions for processing personal data and the latter defines the requirements data logging requirements for communications providers. Both laws are based on the relevant EU directives.

The most important acts in Finland, the Personal Data Act and Act on the Protection of Privacy in Electronic Communications, are based on directives to which these acts are ultimately reflected. The Finnish DPA also refers to the Constitution of Finland, as it defines that protection of personal data must always be regulated and that activities of public authorities must always be based on law. The question is therefore whether Statistics Finland has the right to receive such data and to engage in studies such as this one, although the DPA agrees that most likely this is not an issue. The topic matter of this study is in one DPAs opinion 'faced with constitutional level issues for example in terms of freedom of movement, which is a constitutional right'. Other related constitutional issues include the freedom of assembly. The Act on the Protection of Privacy in Electronic Communications defines for which purposes the mobile network operators are allowed to process telecommunications data.

The granting of licenses will be harmonised by the new European regulation on data protection. The regulation states that if a certain procedure has impact on the residents of other countries or if it incorporates the use of surveillance or monitoring technologies then the evaluation of this procedure will be done on the European level. This is called 'the consistency mechanism'. In one of the opinions, this project faces a risk that the implementation of this mechanism may have an impact. It could also be a schedule risk in the sense that any conclusions made concerning data protection may have to be changed once the consistency mechanism becomes effective.

Overall Opinion of the Usage of Mobile Positioning Data

The overall opinion of the usage of the mobile data is rather good and positive. The respondents see the potential value of the data and consider that it will be used as a source for various statistics in the future given that numerous obstacles can and will be solved.

Respondents have mentioned the growth of the efforts put in investigating new data sources, especially the big data which mobile data is a part of.

This data could really improve knowledge of these mobility flows but it requires a real dialogue between private operator, statistics institutes and the whole society. It is considered a booming market that will soon be out of control of any institution and even the customer segment.

Some of the positive aspects of the usage of mobile positioning data are:

- Great quantity of data available 'without cost';
- Availability of data on intra urban scale;
- Real-time data (timeliness);
- Longitudinal studies would be possible;
- Possibility to cross mobile positioning data with mobile survey.

An opinion was also voice that the data is supplementary to official statistics, but cannot replace them. Someday in future, it could replace the current statistics, but at this moment the methodology is not as sophisticated. The main problem is that currently only one or few mobile operators in one country offer the data, even if there are several other big operators with market share of more than 50%. The usage of mobile positioning data is said to be a great opportunity if correctly used but risky when targeted towards spying citizens. As always this is a deal to take the best and prevent the worse – so it has to be treated as a social issue by ethical committees more than by legal officers, the last are more formally oriented and prefer to be conservative. But it was also pointed out that there should be no issues regarding personal data if the data is made anonymous by the operator prior to the processing for statistical purposes.

It was mentioned that it is important that the statistical community work to establish a right of access to these sources in principle for Official Statistics purposes. It was also pointed out that coordinated and harmonised data exchange via Eurostat would reduce costs and improve quality and comparability. Roaming data would be a powerful supplement to survey data, especially to obtain high quality lower bounds for long distance travel.

Since the amount of information is very large it becomes necessary to set common rules or definitions to guarantee the better harmonisation of the output and construct existing experience between data providers and users. It would also be useful, if there is a website to help with procedures, privacy and other issues, to spread the knowledge, provide references of

operators who grant access to data, along with information on the type of contract and method of transfer used. A so-called knowledge bank describing the process 'from black box to opportunity.'

If it is possible to overcome obstacles and avoid pitfalls, the usage of mobile positioning data could systematically help improve the compilation of tourism data. Even if the stakeholders of possible users of mobile positioning data do not yet have precise plans for using such data, they tend to consider the use of mobile positioning data as a very promising avenue that must be explored seriously for the future of travel statistics.

Although there were optimistic and positive responses, some participants also expressed scepticism towards the methodology, aspects of potential privacy intrusion and infringement of privacy of the subscribers. Most common issues addressed in the methodology were concerning problems arising from different phone usage patterns (some have no phones, some have several phones) and the inability to fully represent the actual phenomena of tourism. Some respondents stated that even if the data is used, it can only supplement and not replace any existing methodologies.

Concerning the privacy aspects, some respondents expressed concern that any kind of indication of the usage of the mobile data might be considered as serious intrusion of the privacy of the subscribers. Even if the personal data is processed anonymously and according to the legislation, the process itself can be perceived as illegal by public.

Conclusions: Data Availability and Accessibility

The opportunities provided by the usage of mobile positioning data are quite widely known already among the stakeholders sampled in the survey (86% of the respondents are aware of the possibilities offered by using mobile positioning data), but it is not used so often yet (mere 14% use it in some form). The data is mainly used by research institutes and private enterprises in their everyday activities as well as for research purposes. Those among survey respondents who have approached mobile operators or other organisations for the possibility of obtaining this type of data, have done it mostly for the purpose of generating tourism statistics. Other important fields that were mentioned are geomarketing, traffic monitoring urban planning. Most commonly, the use of mobile positioning data allows conducting better analysis of tourism flows and mobility, and obtaining the number of tourists/visitors (to tourism areas, events, festivals etc.). Mobile positioning data also helps to analyse tourism mobility, evaluate the number of (inbound) trips and possibly assess the amount of concurrent

costs in some cases. Mobile positioning data is considered to be a valuable complementary tool to other research methods like surveys and obtaining data from other devices, since the current tools tend to be of limited efficiency and a renewed toolset would reduce respondent burden, improve timeliness and data quality.

Despite the low percentage of respondents using mobile positioning data, a relatively large portion of survey respondents (64%) and interviewees are interested in using it, which greatly extends the imperative to make this type of data more available. There were only 8 respondents to the survey who did not see the value in using mobile positioning data in their organisation's processes.

Using mobile positioning data is connected with a high level of secrecy, which is why 35% of the respondents are not allowed to name the mobile network operators who are providing them with the necessary input. Usually only one mobile network operator provides the data through one-time projects. Respondents had to sign agreements or contracts and purchase licenses and specific software to obtain this type of data.

The reasons why respondents have not yet been able to access mobile positioning data are: the process is incomplete, data is incomplete, specific procedures have to be completed and different regulatory issues to be taken into account. Also the financial questions were mentioned.

As a conclusion, based on the survey and interviews, it can be said that there is high and rising interest in the usage of or getting access to mobile positioning data from the user side. Mobile positioning data is irreplaceable and the value it can bring cannot be underestimated. Most often, data in some form is available to organisations and accessible if needed and asked. But at the same time there are also limitations and constraints: the problems, obstacles, hidden pitfalls and risks that have been raised when organisations, research institutes, private enterprises, municipalities and governments express desire to access and use mobile positioning data for their work, research and analysis.

Annex 10. The Initial Responses of the DPA in Finland

Data Protection Ombudsman's Statement on use of CDR not including identification in Finland, from October 4, 2013

THE OFFICE OF THE DATA PROTECTION OMBUDSMAN Ref. 1865/03/2013 4 October 2013

Statistics Finland

00022 STATISTICS FINLAND

With reference to your letter 'Request for statement'
received by the Office of the Data Protection Ombudsman on 20 June 2013

Subject

A request for statement is pending at the Office of the Data Protection Ombudsman, in which Statistics Finland requests from the Office of the Data Protection Ombudsman a statement relating to the use of CDR (call detail records) not including identification data from Finland's mobile telephone operators (Sonera, Elisa and DNA) for tourism statistics. The main data calculated from the data set are the number of trips from abroad to Finland and from Finland abroad and the durations of trips.

According to the request for statement, it is intended that three to four variables, depending on the statistical sub-area, are collected from mobile sub-scriber connections. The subscriber connection's identifying code, time-stamp and rough location data would be collected in all sub-areas. The sub-scriber connection's identifying code is 'An artificial code generated for the selection, by which an individual subscriber connection is distinguished from other subscriber connections inside the data. The subscriber connection cannot with the help of the code be combined with other data to find out the identity of the owner of the data.' The timestamp is in turn 'The timestamp of an activity (call / SMS) made from the subscriber connection (date and time).' Rough location data are 'The rough location data of the mobile net-work antenna having transmitted the activity, such as a municipality, postal code or other corresponding administrative area. Location data enable compilation of statistics on the regional direction of tourist flows inside Finland.' In addition to this, the home country of the subscriber connection or mobile network is collected in some cases as well.

In its request for statement, Statistics Finland has inquired, whether the use of the data described by it for Statistics Finland's tourism statistics can be considered anonymous processing of location data. In addition, Statistics Finland has asked, whether a telecommunications operator can process the data needed in the survey by virtue of Section 12a of the Act on the Protection of Privacy in Electronic Communications or on some other possible basis and release the data requested to Statistics Finland, and whether it is relevant in this case by whom and at what stage the artificial code is created. According Section 12a of the Act on the Protection of Privacy in

Electronic Communications, a telecommunications operator may process identification data for statistical analysis under certain conditions.

Owing to the pending matter, further clarification was requested on the matter (telephone conversation Leivonen/Office of the Data Protection Ombudsman - Nurmi/Statistics Finland on 2 Oct. 2013). During the telephone conversation, it was asked for which statistics the data in question are specifically needed. According to Nurmi, data were to be utilised for tourism statistics. Statistics Finland produces data on tourism income and expenditure for the Bank of Finland, for example. Statistics are intended to be made on tourist volumes, that is, on how many tourists come to Finland and how many leave Finland. On the telephone, Nurmi was asked for what purpose the timestamp is needed (call/SMS data). Nurmi said that that type of data are meant to be collected based on Estonia's reference model. The duration of one trip could be deduced based on timestamps. Timestamps can also be used to monitor the trip and there is a need for statistics both on the national level and for travel inside Finland. For example, previously tourism could be examined based only on overnight stays, this proposed model would allow statistics to be compiled on same-day trips as well. According to the clarification, the collection of data is not meant to be restricted by time.

Competence of the Data Protection Ombudsman

According to Section 38, Paragraph 1 of the Personal Data Act, the Data Protection Ombudsman provides direction and guidance on the processing of personal data and supervises the processing in order to achieve the objectives of this Act, as well as makes decisions, as provided in this Act.

According to Section 40, Paragraph 1 of the Personal Data Act, the Data Protection Ombudsman shall promote good processing practice and issue directions and guidelines so as to achieve a situation where unlawful conduct is not continued or repeated. Where necessary, the Data Protection Ombudsman shall refer the matter to be dealt with by the Data Protection Board, or report it for prosecution.

According Section 32 of the Act on the Protection of Privacy in Electronic Communications, the duties of the Data Protection Ombudsman include the supervision of the processing of location data in accordance with Chapter 4 of the Personal Data Act.

Reply

Chapter 4 of the Act on the Protection of Privacy in Electronic Communications provides for the processing of location data. According to Section 2, Paragraph 9, Sub-paragraph 9 of the Act on the Protection of Privacy in Electronic Communications, location data refer to data which show the geo-graphic location of a subscriber connection or terminal device and which are used for a purpose other than the provision of a network service or communications service. Location data may indicate the latitude, longitude and height of the user's terminal device, the direction of the trip, the accuracy of location data, the part of the network where the terminal device is located at a specific moment, and the time of storing the location data. Some data, such as data in the mobile network data showing in which base station area the active mobile transmitter is at a certain moment, are included in the concept of identification data, because these data are used for transmitting messages. If base station data are used for other purposes than transmitting messages, however, they are not necessarily identification data intended in this Act but location data.

According to Section 16 of the Act on the Protection of Privacy in Electronic Communications, 'Telecommunications operators, value added ser-vice providers and corporate or association subscribers and any persons acting on their behalf may process location data subject to the provisions of this Chapter for the purpose of providing and using value added services.' Ac-cording to Section 2, Paragraph 1, Sub-paragraph 7, value added services re-fer to a service based on the processing of identification data or location data for a purpose other than the provision of a network service or communications service. The Government proposal concerning the Act on the Protec-tion of Privacy in Electronic Communications states that 'Value added ser-vices can be advertising and route information, traffic information, weather forecasts or tourism information based on the location of the user's terminal device.'

According to Section 17 of the Act on the Protection of Privacy in Electronic Communications, 'A telecommunications operator may process loca-tion data if the subscriber has not forbidden it.' If a telecommunications op-erator wishes to release location data to a value added service provider or corporate or association subscriber, the telecommunications operator must ensure appropriately before releasing the location data that the provision of the value added service is based on the consent referred to in Section 18, Paragraph 1.

According to Section 16 of the Act on the Protection of Privacy in Electronic Communications, the provisions of Chapter 4 of the Act on the Protection of Privacy in Electronic Communications are not applied to location data if they are rendered such that they cannot, in themselves or in combination with other data, be associated with a specific subscriber or user, unless otherwise provided by law. In accordance with Section 3 of the Act on the Protection of Privacy in Electronic Communications, a subscriber refers to a legal person or a natural person who uses the message service or value added service without necessarily being a subscriber to that service. If location data can be associated either with the subscriber or user, it can also be a question of personal data according to Section 3, Paragraph 1, Sub-paragraph 1 of the Personal Data Act. Location and locating data can be regarded as personal data according to the Personal Data Act when a natural person can be identified directly or indirectly from the data.

The essential question is whether location data planned to be released to Statistics Finland as described in the request for statement are rendered such that they cannot, in themselves or in combination with other data, be associated with a specific subscriber or user.

If location data are such that they fulfil the definition of personal data, it may also be considered that location data may then be associated with either a subscriber or user. Therefore, the question presented can be approached through the concept of personal data. According to Section 3 of the Personal Data Act, personal data means any information on a private individual and any information on his/her personal characteristics or personal circumstances, where these are identifiable as concerning him/her or the members of his/her family or household. The Data Protection Working Party established by Article 29 of the Directive on the protection of personal data has stated in its opinion of 4/2007 on the concept of personal data concerning indirect identification: 'If the available identifiers do not, as a rule, enable identification of a person, the person may still be 'identifiable' because the data combined with other data (which need not be in the possession of the controller) enable distinguishing the person from others.

It is important to notice that mere removal of names and other individualising data does not in all cases make all data anonymous. The collected data may also contain detailed data on the data subject (e.g. enough data types or

observations) so that the data subjects are indirectly identifiable, even if direct personal identifiers are not processed. It is also crucial to observe that identification of a person does not mean only finding out the person's name but the person can be identified through other data as well.

What particularly stands out here is that it is not merely a question of releasing rough location data, but also the subscriber connection's individualising identifier, timestamp and in some cases also data on its home country would be released. In order to be able to separate different trips of a certain person, in my view, this would require that the subscriber connection's individualising identifier would always be the same for a certain mobile phone. So that observations concerning one subscriber connection could be distinguished from among other observations, the identifier individualising the subscriber connection would have to be the same and for this reason, the identifier should be such that it is still identifiable as concerning a certain subscriber connection at least by the telecommunications operator producing the data. If the identifier concerning the subscriber connection were varying, this in-formation would not to my understanding serve the compilation of tourism statistics intended now in the request for statement, because data on a certain trip could be partial or otherwise fragmental. In connection with the clarification received, no restrictions are set for the collection time of observations. Thus, examining tourism inside Finland, for instance, would require continuous collection of data for this purpose. The number of observations to be collected is not limited either according to the clarification received. It has also not been established what tourism means and how tourism would be separated from other movement, for example (e.g. travel to work) in this context. Because no stand is taken on the matters above in the request for statement, the activity described in the request for statement enable collection of data even so that Statistics Finland would end up with a considerable number of observations concerning one subscriber connection. **Even if the location data were made rough** (rough location data of the mobile network antenna, such as a municipality, postal code or other corresponding administrative area) **as presented in the request for statement, it will not alone guarantee anonymity of data if the identifier individualising the sub-scriber connection is the same all the time, plenty of observations are collected from the subscriber connection by means of location data and timestamps, and the time of data collection and the number of observations accumulated in this way are in no manner restricted. For this rea-son, based on the issued request for statement and other clarification received on the matter, it cannot be considered that the planned activity would be 'anonymous processing of location data'.**

In its opinion concerning the concept of personal data, the Data Protection Working Party WP 29 has stated that the principles concerning data protection are not applied to data that are made anonymous so that the data subject is no longer identifiable. According to the Data Protection Working Party, it depends on the situation whether a person can be identified with the help of the data and whether the data can be considered anonymous or not. Each case must be viewed separately by taking into account to what degree reasonably implementable measures are likely to be used to identify the person concerned.

Concerning the request for statement, it should be noted that it is written on a very general level. Evaluation of the anonymity of data calls for case-specific consideration, which requires, in addition to what is now presented in the request for statement, a detailed process description of the life cycle of the data processed in the activity starting from raw data (CDR, etc.) to final data to be released and processed, individualising the data processors for each stage of processing, and information on how much and over what time period the data will be collected.

Release of statistical level data produced from location data can be possible when the data are rendered such that location data in themselves or in combination with other data cannot be associated with a subscriber or user. This also requires that statistical level data are anonymised and aggregated so that for personal data, the possibility for indirect identification is precluded. Evaluating what level of aggregation and what level of anonymising techniques will preclude the possibility for indirect identification is a challenging task. Therefore, data protection provisions have an important role in analysing what kinds of data could be regarded anonymous so that the conditions set by the data protection provisions need not be taken into account in their processing. In the end, the controller is responsible for the lawfulness of the processing of personal data.

The provisions concerning the protection of location data and personal data shall be applied as long as the data in question can be restored to those with identifiers or otherwise indirectly identified as relating to a certain person.

Finally, take notice of the fact that the Data Protection Ombudsman does not have power of acceptance or power to grant permission.

To the respect that your question regards the applicability of Section 12a of the Act on the Protection of Privacy in Electronic Communications, the competent authority is the Finnish Communications Regulatory Authority (see Section 31 of the Act on the Protection of Privacy in Electronic Communications).

Data Protection Ombudsman Reijo Aarnio

Senior Officer Raisa Levonen

For the information of: Ossi Nurmi, ossi.nurmi@stat.fi

Competence of the Data Protection Ombudsman

According to Section 38, Paragraph 1 of the Personal Data Act (523/1999), the Data Protection Ombudsman provides direction and guidance on the processing of personal data, supervises the processing in order to achieve the objectives of this Act, as well as makes decisions, as provided in this Act.

According to Section 40, Paragraph 1 of the Personal Data Act, the Data Protection Ombudsman shall promote good processing practice and issue directions and guidelines so as to achieve a situation where unlawful conduct is not continued or repeated. Where necessary, the Data Protection Ombudsman shall refer the matter to be dealt with by the Data Protection Board, or report it for prosecution.

According to Section 40, Paragraph 1 of the Personal Data Act, the Data Protection Ombudsman shall decide matters brought to his/her attention by data subjects on the basis of Sections 28 and 29. The Ombudsman may order a controller to realise the right of access of the data subject or to rectify an error.

The Data Protection Ombudsman supervises the compliance with Section 22 of the Act on the Protection of Privacy in Working Life (759/2004) together with occupational health and safety authorities.

According to Section 32 of the Act on the Protection of Privacy in Electronic Communications, the Data Ombudsman shall supervise compliance with the provisions on direct marketing in Chapter 7 (Sections 26 to 29).

Interview with Reijo Aarno, the Data Protection Ombudsman, Finland, on October 4, 2013

A) To your knowledge, are mobile positioning data already being used for one or more of following purposes:

1. … anonymized or non-anonymized identification or tracing of individual mobile devices for public or private purposes?

   We are aware of the pilot study conducted by Finnish Road Administration concerning the measurement of average speed in certain road sections. We are not aware of other studies based on anonymous mobile positioning data. When discussing this type of studies my advice for your project is to make sure that no classified information is disclose when reporting results. I noticed that in your questionnaire there were certain questions concerning business secrets.

   The case concerning Google Streetview was also collecting data on the locations of mobile base stations. Due to a certain software bug also communications signals were collected.

   It is highly probable that anonymized location data of the mobile phone is currently being used for monitoring purposes. For example base stations collect statistical information when the mobile device moves from one base station to the other. This data is not necessarily used for statistical purposes but rather for the enabling of communications. This is a case concerning the processing of telecommunications data. The Finnish legislation clearly states that the processing of positioning data requires consent of the subscriber. Certain applications / solutions have been made in this area but usually the legal basis for these applications has been consent of the subscriber.

2. ... applications like traffic monitoring (FCD), police or emergency services etc?

   According to the Act on the Protection of Privacy in Electronic Communications the police and emergency response centres are entitled to access the positioning data to the extent where they are considered telecommunications data. Positioning data in this context is data stored for a purpose other than transmitting communications.

3. …more specifically: e.g. for the enhancement of official statistics?

   As far as I'm aware, mobile positioning has not been used for the production of official statistics. (Statistics Finland confirms that this is true).

4. Have you been involved in the process of using mobile positioning data in the aspect of assuring privacy protection of the subscribers?

It must be stated that the Office of Data Protection Ombudsman has no legal power to grant or prevent licenses to such projects / applications. I find it a bit strange that the mobile operators aspire for a kind 'blessing' of the Data Protection Ombudsman for the project. It's not a good starting point if the operators, who have a central role in the project, are not themselves exactly aware what the legislation says and request to consult a supervising authority in this matter. In my opinion, the operators should be able to state what they are allowed to do considering the legislation. Like I said, the Data Protection Ombudsman doesn't possess such authority to grant licenses or permits.

B) If yes: What do you know about these projects or applications? Please give more detailed information on each of the projects or applications (or name a reference):

1. Which companies/organisations were involved?

   Finnish Road Administration (currently Finnish Transport Agency) and former Radiolinja (currently Elisa).

2. What data were used? How were they retrieved?

   The method is based on the time it takes for a cell phone to switch from one base station to another in the at predetermined observation points in the antenna network.

3. Was/Is it a onetime project or a continuous operation? Is it still active?

   This was a one-off pilot project conducted in 2002. We are not aware of continuation of this project.

4. What were the main learnings? Would you work with the same setup again? What would you change?

5. Is there any documentation on the project? Is it accessible?

(The request and statement related to the pilot project of Finnish Road Administration have been provided earlier.)

C) What are, in your opinion, main barriers for the implementation of a positioning service for tourism statistics? Do you see any solutions? Please consider the following aspects.

1. Privacy and regulatory issues:
   a. What are the common procedures (concerning your agency in your country) when mobile network operators want to provide any kind of (aggregated or anonymised raw) mobile positioning data?

There is no generic answer for what is processing of truly anonymous positioning data. In one of the documents related to the preparing of the Act on the Protection of Privacy in Electronic Communications, there is an example of this. In this example it is stated that if anonymous positioning data is combined to video surveillance data, the combined data immediately becomes personal data. The European Union is currently preparing a document that hopefully will give guidelines to what anonymisation means in practise. Generally speaking, if data is anonymized it does not concern the Data Protection Ombudsman. However for the operator even the anonymized data is still personal data as the operator knows the identity of the subscriber.

As far as geo data is concerned, the possibility to indirectly identify the subscriber depends on the size of the geographical grid used. The data can be personal data if the grid is too small.

Each operator knows the number of roaming clients given a certain timeframe. For example, the operators could certainly report statistics on the number of roaming subscribers separately for each nationality along with duration of visit based on the roaming time. If for example an operator couldl report that we had 100 000 Swedish roaming clients this quarter. Such statistics would not be considered personal data and the operator is certainly capable of providing this. The question whether Statistics Finland is entitled to collect such data is a separate question and is defined in the Statistics Act.

Often the mobile positioning data is also telecommunication data. The processing of telecommunications in Finland is supervised by the Finnish Communications Regulatory Agency. They are also more capable of commenting whether telecommunications data can be processed for your purposes.

b. How is the privacy of the subscribers guaranteed?

Considering the Finnish legislation a principle requirement called 'privacy by design' has been in place for 25 years. Considering our national statistics this principle means that there are most likely several alternatives in addition to the model implemented in Estonia. These alternatives should also be considered. To mention a few example, there are GPS-based solutions, traffic / passenger counts at airports and ports. The EU is also implementing an EES-system (entry and exit) to collect data on passengers leaving and entering the Schengen area. This would be perfect data source for non-Schengen visitors but obviously doesn't measure tourism flows within the Schengen area.

Another aspect is the principle of minimisation: is less sufficient? Does the statistics have to be based on exhaustive data or is a sample sufficient?

I also want to present a general question: what is meant by data protection? According to the recital of the data protection directive, data protection systems are meant to serve people and not the opposite. When systems are developed, the

fundamental rights of people are to be taken into account, especially the right for privacy. The concept of privacy is not defined any further. Fundamentally, data protection means that every person has the right to decide. I decide whether I want to be included in Statistics or not. I decide whether I keep my phone open or not. I decide whether I have a prepaid, confidential or public subscription etc.

c. What legislation applies in your country concerning the usage of mobile positioning data?

The most important acts, the Personal Data Act and Act on the Protection of Privacy in Electronic Communications, are based on directives to which these acts are ultimately reflected. Also, when considering personal data and public authorities, one must not forget the Constitution of Finland. It defines that protection of personal data must always be regulated and that activities of public authorities must always be based on law. The question is therefore whether Statistics Finland has the right to receive such data and to engage in studies such as this one. Most likely this is not an issue. Considering this study, we are in my opinion faced with constitutional level issues for example in terms of freedom of movement, which is a constitutional right. Other related constitutional issues are freedom of assembly. Considering these constitutional freedoms, I personally consider it radical if Finnish authorities are monitoring the movement of people for the sake of promoting tourism.

The Act on the Protection of Privacy in Electronic Communications defines for which purposes the mobile network operators are allowed to process telecommunications data. I still would like to stress that processing of telecommunications data is supervised by the Finnish Telecommunications Regulatory authority.

The granting of licenses will be harmonized by the new European regulation on data protection. The regulation states that if a certain procedure has impact on the residents of other countries or if it incorporates the use of surveillance or monitoring technologies then the evaluation of this procedure will be done on the European level. This is called 'the consistency mechanism'. In my opinion, your project faces a risk that the implementation of this mechanism may come as a surprise. It could also be a schedule risk in the sense that any conclusions made concerning data protection may have to be changed once the consistency mechanism becomes effective.

2. Financial and business issues:

3. Technological issues:

D) Are there any advices to the project consortium?

1. Obstacles, hidden pitfalls, challenges, risks:

The risk I already mentioned concerning the upcoming European data protection regulation and the consistency mechanism.

One of the major challenges in this kind of endeavours is the challenge of communication and transparency. Do people wish that data about their movements can be accessed by public authorities?

According to European studies the Finns are exceptional in their confidence in public authorities. For example, the studies outline that 92% of Finnish citizens trust the police. Considering this background I think you are faced with a huge challenge in how to communicate the need for this type of monitoring for the common people.

Other questions that your project needs to tackle are the special cases. In the Act on the Protection of Privacy in Electronic Communications there is a section on the special rights to receive data. For example, the police is entitled to receive data and the person himself is entitled to receive data of his own movements. Let's consider a case where a company wants to validate a travel invoice made by an employee. My experience is that once systems are implemented there will always be a queue of requests for that data.

Considering the operators, I would like to point out that the choices made in the technical implementation of a service may also have an impact concerning the legislation.

2. Helpful partners, opportunities:

   You should contact the Finnish Telecommunications Regulatory Agency since the positioning data in this context is also telecommunications data.

3. What is your overall opinion about the usage of mobile positioning data concerning all the positive and negative aspects?

   I see this topic as a constitutional issue related to the freedom of movement. A common person need not be under continuous surveillance unlike for example a professional athlete. This type of surveillance model would challenge this constitutional right which might be a significant issue.

   Another issue the challenges we have with anonymisation skills and knowledge. Finland is a land of registers and we are currently in the process of opening these registers for public use free of charge. Doing so, we should also have high quality skills and knowledge in data anonymisation and planning.

   From the perspective of the operator, the data in question is without doubt not anonymous. The operator is fully aware of the identity of the subscriber. Does the operator then have the right to process this data for this purpose? According to the Act on the Protection of Privacy in Electronic Communications, even the process of making the data anonymous is considered processing of telecommunications data. The act defines the purposes for which the data can be processed. It cannot be clearly pointed out based on this act that there would be a legal basis for the processing of

telecommunications data for this kind of purpose. Also there does not seem to be a legal right or obligation for the operators to submit this data for example to Statistics Finland.

Prior to the coming into force of the new European data protection regulation my concluding analysis is the following: the current legislation would most likely need to be updated to enable the type of data collection as you have presented in your study. More information can be found in our statement that we have delivered to Statistics Finland today.