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

Eurostat contract no. 30501.2012.001-2012.452
Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

The Consortium:

Statistics Finland

Eesti Statistika

NIT

UNIVERSITY OF TARTU

IFSTTAR

positium

Project time: January 2013 - March 2014
Project website: mobfs.positium.ee
The Aim of the Project

Exploring the possibilities - and limits - of using mobile positioning data stored by mobile network providers for generating tourism statistics
Main project objectives

• Assess feasibility to access databases with mobile positioning data in European countries
• Assess the feasibility to use mobile positioning data for tourism statistics in the European context
• Identify, discuss and address the main challenges for implementation
• Assess the potential impact on cost-efficiency of data production
• Assess the possibility to expand the methodology to other domains and define joint algorithms
Main project objectives

• Mix of scientific/theoretical & practical/empirical/applied work

• Can the technology/methodology be applied to the particular case of tourism statistics, across a wide group of countries in a similar way? Can the outcomes be generalised to all countries?

• The work already done is the starting point for the multidisciplinary consortium, not an intermediate result to be recycled in voluminous reports
Majority of the people use mobile phones on daily basis...

...leaving geographical footprint in the databases of Mobile Network Operators (MNO)

Processing the historical data with the aim of producing collective statistics...

...results in statistical indicators that help understand the mobility of people, e.g. in tourism
Project tasks

Task 1: Stock-taking
Task 2: Feasibility of Access
Task 3a: Methodology
Task 3b: Coherence
Task 4: Opportunities and Benefits
Task 5: Visibility and Consolidated Report
TASK 1: STOCK-TAKING
Objective:
to map the use cases of mobile positioning data in Europe:

• In tourism statistics

• Other statistical products
  • Tourism studies

• Scientific research and other projects
Documented 31 significant and representative cases with access to data.

Data initiatives started 2002 in Estonia with EMT / Telia & Sonera MPS tracking.
Mobile data for R/D

- **2002** - Estonia - MPS tracking in urban studies, University of Tartu
- **April 2004** - Estonia - CDR data collection, Positium LBS
- **2005** - Austria - „Graz in real time“, MIT Sensible City Lab
- **2006** - Portugal - „Socio-Geography of Human Mobility“, Orange Lab
- **2006** - Italy - „Rome in Real Time“, MIT Sensible City Lab
- **2009** - France - „Paris Tourism with CDR“, Orange lab
- **2009** - Ireland - "Utilising Mobile Phone RSSI Metric.. “University of Ireland Maynooth, IBM Research“
- **2009** - Switzerland - „Mobile Data Challenge“, Nokia
- **2010** - Czech Republic - CE Traffic, traffic analysis
- **2012...** Telefonica, Orange - commercial offerings
Mobile data for tourism statistics

2008 - Estonia - Central Bank started to use mobile data for „Balance of payment calculation“ Positium LBS

2010 - the Netherlands - „Time patterns, geospatial clustering“ Statistics Netherlands

2012 - Czech Republic - Czech Tourism

Three types of cooperation:

1. Short-term projects

2. Direct business solutions, some with clear profit models

F1 success in 1990: budget, car, driver...
F1 success in 2014: 
budget, sensors, ...
Stock-taking conclusions II:

Need for fast data

Need for automatic data collection

In 2-5 years mobile data will probably become a common source for studies and statistics

Businesses and Location Based Services will follow ...
Questions?
TASK 2: FEASIBILITY OF ACCESS
Feasibility of Access

- Privacy and Regulations (EU & national)
- Technology
- Financial and Business barriers
- Practical Experience on Accessing the Data

Also during the Task:
- Online survey
- Interviews with stakeholders
- Legal analysis
Online Survey & Interviews

Learn the state of mind of the stakeholders

Understand specific barriers and solutions in gaining access to the data

Understanding the experience concerning privacy protection, regulatory, business, financial and technological aspects
Online Survey & Interviews

Survey:
• Sent to 422 respondents
• 118 responses
• 115 organisations
• 35 countries (16 EU)

37 interviews:
• 7 DPAs
• 11 MNOs
• 5 NSIs
• 4 research institutes
• 1 central bank
• 9 public organisations for tourism
Awareness

Respondents

- Aware: 86%
- Interested: 64%
- Have approached: 17%
- Some experience: 14%
- Not interested: 9%
Expectations

Better temporal and spatial accuracy

New statistical indicators

Volumes of travellers, event visitors

Duration of trips

Travel routes

Point of entry

Places visited

Plausibility checks of tourism data

Faster data generation

Reduced respondent burden
MNOs

Mostly sympathetic to the idea
Concerns regarding legal restriction and obligations to provide the data
Public opinion and a possible decline in the number of clients due to bad reputation
Value for the MNOs if they provide the data
Main Obstacles of Accessing the Data

- Privacy concerns: 62%
- Regulatory and legislation obstacles: 62%
- High implementation and maintenance cost: 32%
- Technological issues: 27%
- Exposure of business secrets: 23%
- Other: 13%
- No business value: 5%
- Conceptual issues and meaningfulness of data: 1%
Privacy Protection (EU Directives)

The first and main ‘barrier’ for accessing the data

Regulations governing the subject:

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

Directly or indirectly identifiable (mobile positioning) data can be used and processed for statistics if the following is true:

1. The subscriber has given their consent, or
2. National legislation allows the NSI and compels the MNOs, or
3. For the purpose of historical or scientific research
Privacy Protection

Fully anonymous (mobile positioning) data can be processed and used without restriction given that subscribers cannot be directly or indirectly identifiable at any given processing stage.

Fully anonymous (currently) = no longevity
Privacy Protection: Grey Area

Because the end result of processing is, by itself, anonymous (involving aggregated number, duration, and the travel patterns for collective trips),

the processing of personal data for such purpose can be interpreted as appropriate
Legislation: NSI Rights to data

Different rights of access for NSIs:

1. Legislation-based (new data source requires possible amendments to the legislation, more difficult and time-consuming to implement)

2. Directive-based (requiring a simple addition to the list of sources for national official statistics)

3. Unconditionally applied (involving the power of NSIs to request valuable data from businesses)

NSI should verify their position in relation to their Statistical Act, defining their right to access a new data source - access to mobile data can range from easy to nearly impossible
Privacy Protection

No single clear understanding on how mobile data can be accessed

Differences in countries: Estonia vs Finland vs France vs Germany ...

Suggestion for a unified legal/methodological framework for NSIs to access the data from MNOs
Technical Accessibility

Source of the data
Initial processing by MNOs
Allocation of data compilation processes
Technological limitations
Continuity of access
Forms of tourism activity is broken down the same way as MNO internal databases, although they do not correspond in every case (Inbound roaming subscribers can be a part of domestic tourism if they are classified as residents based on the duration of trips within the country.)
<table>
<thead>
<tr>
<th>Subscriber</th>
<th>Activity</th>
<th>Time</th>
<th>Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>244211964246</td>
<td>Data</td>
<td>07.04.2014 12:15:00</td>
<td>43879</td>
</tr>
<tr>
<td>244121965188</td>
<td>LU</td>
<td>07.04.2014 12:15:02</td>
<td>43879</td>
</tr>
<tr>
<td>206201963365</td>
<td>Call</td>
<td>07.04.2014 12:15:04</td>
<td>44866</td>
</tr>
<tr>
<td>244121965188</td>
<td>Call</td>
<td>07.04.2014 12:15:04</td>
<td>43879</td>
</tr>
<tr>
<td>244121965188</td>
<td>Call</td>
<td>07.04.2014 12:15:05</td>
<td>43879</td>
</tr>
<tr>
<td>244211964246</td>
<td>LU</td>
<td>07.04.2014 12:15:06</td>
<td>43877</td>
</tr>
<tr>
<td>244121965188</td>
<td>Data</td>
<td>07.04.2014 12:15:09</td>
<td>43879</td>
</tr>
<tr>
<td>24405239944</td>
<td>LU</td>
<td>07.04.2014 12:15:09</td>
<td>48512</td>
</tr>
<tr>
<td>244211548784</td>
<td>Data</td>
<td>07.04.2014 12:15:10</td>
<td>48987</td>
</tr>
<tr>
<td>244121964444</td>
<td>SMS</td>
<td>07.04.2014 12:15:12</td>
<td>45559</td>
</tr>
<tr>
<td>244051604891</td>
<td>Data</td>
<td>07.04.2014 12:15:15</td>
<td>45601</td>
</tr>
<tr>
<td>24201725641</td>
<td>LU</td>
<td>07.04.2014 12:15:18</td>
<td>45463</td>
</tr>
<tr>
<td>244051965315</td>
<td>Data</td>
<td>07.04.2014 12:15:20</td>
<td>48987</td>
</tr>
<tr>
<td>244211963912</td>
<td>Call</td>
<td>07.04.2014 12:15:20</td>
<td>43570</td>
</tr>
<tr>
<td>244051605773</td>
<td>SMS</td>
<td>07.04.2014 12:15:23</td>
<td>35550</td>
</tr>
<tr>
<td>244211914278</td>
<td>Data</td>
<td>07.04.2014 12:15:26</td>
<td>48987</td>
</tr>
<tr>
<td>24421417297</td>
<td>Call</td>
<td>07.04.2014 12:15:26</td>
<td>48987</td>
</tr>
<tr>
<td>24421838967</td>
<td>LU</td>
<td>07.04.2014 12:15:27</td>
<td>43951</td>
</tr>
<tr>
<td>244051965316</td>
<td>Data</td>
<td>07.04.2014 12:15:29</td>
<td>43909</td>
</tr>
<tr>
<td>244211964762</td>
<td>LU</td>
<td>07.04.2014 12:15:30</td>
<td>46844</td>
</tr>
<tr>
<td>23802964803</td>
<td>Data</td>
<td>07.04.2014 12:15:30</td>
<td>46424</td>
</tr>
<tr>
<td>24412179162</td>
<td>LU</td>
<td>07.04.2014 12:15:30</td>
<td>46977</td>
</tr>
<tr>
<td>24421768224</td>
<td>Data</td>
<td>07.04.2014 12:15:32</td>
<td>43878</td>
</tr>
<tr>
<td>260021793995</td>
<td>LU</td>
<td>07.04.2014 12:15:35</td>
<td>43880</td>
</tr>
<tr>
<td>24405447807</td>
<td>Data</td>
<td>07.04.2014 12:15:37</td>
<td>38867</td>
</tr>
<tr>
<td>21670369602</td>
<td>SMS</td>
<td>07.04.2014 12:15:40</td>
<td>48302</td>
</tr>
<tr>
<td>22603714049</td>
<td>Data</td>
<td>07.04.2014 12:15:42</td>
<td>43878</td>
</tr>
<tr>
<td>23411962580</td>
<td>SMS</td>
<td>07.04.2014 12:15:44</td>
<td>43948</td>
</tr>
<tr>
<td>20081819893</td>
<td>SMS</td>
<td>07.04.2014 12:15:44</td>
<td>45464</td>
</tr>
<tr>
<td>244051160625</td>
<td>LU</td>
<td>07.04.2014 12:15:44</td>
<td>48987</td>
</tr>
<tr>
<td>244121965280</td>
<td>SMS</td>
<td>07.04.2014 12:15:46</td>
<td>7430</td>
</tr>
</tbody>
</table>
## Types of Data

<table>
<thead>
<tr>
<th>Data extraction option</th>
<th>Number of events per subscriber per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call Detail Records (CDR)</td>
<td>~3…8</td>
</tr>
<tr>
<td>Data Detail Records (DDR, internet usage)</td>
<td>~20…300</td>
</tr>
<tr>
<td>Location updates</td>
<td>~15…35</td>
</tr>
<tr>
<td>Technical data</td>
<td>~25…100</td>
</tr>
<tr>
<td>BTS/NodeB/eNodeB and BSC/RNC/MME (Abis level)</td>
<td></td>
</tr>
</tbody>
</table>
Initial Event Data

Inbound roaming

- subscriber_id
- country_code
- event_time
- event_type
- cell_id

Outbound roaming

- subscriber_id
- event_time
- event_type
- country_code

Domestic data

- subscriber_id
- event_time
- event_type
- cell_id

Some socio-demographic data
Location of the Antennae

<table>
<thead>
<tr>
<th>subscriber_id</th>
<th>cell_id</th>
<th>event_time</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>34256</td>
<td>2013-06-12 14:30:12</td>
<td>8479654</td>
<td>356482153</td>
</tr>
</tbody>
</table>

[Diagram showing the location of the antennae and the relationship between the phone's location and the cell tower data.]
Processed by MNO

1. Removal of non-representative data
2. Black-list (and sampling)
3. Formatting
4. Quality assurance
5. (Pseudonymisation)
6. Delivering the event data to the processor
Allocation of Data Compilation Processes

MNOs process and transmit the result statistical indicators
NSI receives and combines statistical indicators

MNOs extract and transmit the raw data
NSI receives and processes the raw data into statistical indicators
Threats Continuity of Data Access

Major global shift in mobile technology

Changes of the characteristics of the data

Administrative changes (e.g. changed number of providing MNOs)

Can have positive, negative or unforeseen effects on data quality. It is necessary to remain flexible in methodology and estimation to adapt changes in the number of data-providing MNOs.
Technological Limitations

Differences in Network Systems

Data peculiarities

Technological patents

Processing issues (amount of data)

Continuous data update issues (processing time)

Overall, technological capability is not seen as a hard barrier to access
Financial and Business barriers (MNO viewpoint)

Technically, extraction entirely possible

Interested in discussing the potential of the data, if the following is considered:

1. Legal aspects and regulations
2. Public opinion
3. MNO burden high (implementation of extraction system, human resources, maintenance)
4. Business secrets (e.g. share in the country’s roaming market)
5. What are the benefits for MNO?

MNOs expect a mutually beneficial relationship: a) a remuneration scheme or b) being able to use the resulting data themselves for other (including internal and profit-making) purposes
Practical Experience on Accessing the Data

Plan to conduct pilod tests with data in at least 4 different countries

Previous access from Estonian MNOs

Attempts to access data from FI, DE, FR and other MNOs not successful

Available data not usable (initial low value aggregates) and too expensive
Feasibility of Access: Conclusions

High interest from the users
Difficulties to access the data
No clear legal framework to access
Technologically feasible
Need to incentivize MNOs
Questions?
TASK 3A: METHODOLOGY
Data Extraction

From initial raw data to formatted and prepared processable data

Querying data from registries / databases

Geographical referencing and distribution

Removing M2M devices

Extracting the country codes

Formatting the initial data
**Data Extraction**

<table>
<thead>
<tr>
<th>Subscriber</th>
<th>Activity</th>
<th>Time</th>
<th>Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>244211964246</td>
<td>Data</td>
<td>07.04.2014 12:15:00</td>
<td>43879</td>
</tr>
<tr>
<td>244121965188</td>
<td>LU</td>
<td>07.04.2014 12:15:02</td>
<td>43879</td>
</tr>
<tr>
<td>206201963365</td>
<td>Call</td>
<td>07.04.2014 12:15:04</td>
<td>44866</td>
</tr>
<tr>
<td>244121965188</td>
<td>Call</td>
<td>07.04.2014 12:15:04</td>
<td>43879</td>
</tr>
</tbody>
</table>
Probabilistic geographic distribution *

* Required for spatial accuracy on smaller administrative units
Frame Formation

From prepared initial data to formation of the frame population (tourism)

Country of Residence

Place of Residence

Usual Environment

Travels outside Usual Environment
Combined Approach to Country & Place of Residence, and Usual Environment
Identifying Usual Environment

Criteria for measuring routine activities
Cross-border commuting
Limitations due to the lack of data from other countries
Not possible to ask
Large differences due to definitions
Example: Domestic Trips Outside Usual Environment

Using LAU-2 for defining usual environment
Using LAU-1 for defining usual environment
Data Compilation

Assigning statistical indicators

Time aggregation
(hour, day, week, month, quarter, year)

Geographic aggregation
(country, region, municipality, grid)
Estimation and Combining

Reference data
(accommodation, population/census, travels, border survey...)

MNO specific estimation process

Combining estimates from several MNOs

Dissimination of the data
Tourism Statistics Indicators

Breakdown:

Country of residence/place of residence
Aggregation of time (day, week, month)
Aggregation of space (different level of admin. units, grid)
Duration of trip/stay (length, same-day/overnight)
Destination, secondary destination, transit pass-through
Collective movement patterns
Repeat visits

Indicators:

Number of trips/visits
Number of nights spent
Number of days present
Duration of trips
Number of unique visitors

Many indicators coincide with traditional indicators but lacking several classification aspects that are required for tourism statistics.
Anonymous / aggregated initial data

- country of residence
- place of residence
- usual environment
- migration
- trips (geographical routes)
- duration of trips
- repeat visits to places
- distinguish tourism / regular trips

Longitudinal data

- country of residence
- place of residence
- usual environment
- migration
- trips (geographical routes)
- duration of trips
- repeat visits to places
- distinguish tourism / regular trips
Limitations of the data source

No accommodation

Mostly unknown purpose of the trip

No expenditure information

Mostly unknown means of transportation

Usually no socio-demographic breakdown
‘tourism’ means the activity of visitors taking a trip to a main destination outside their usual environment, for less than a year, for any main purpose, including business, leisure or other personal purpose, other than to be employed by a resident entity in the place visited

-Regulation EC 692/2011
Official Indicators: Example 1

‘tourism’ means the activity of visitors taking a trip to a main destination outside their usual environment, for less than a year, for any main purpose, including business, leisure or other personal purpose, other than to be employed by a resident entity in the place visited.

-Regulation EC 692/2011
Official Indicators: Example 2

Arrivals of residents at tourist accommodation establishments

- Regulation EC 692/2011
Official Indicators: Example 2

Arrivals of residents at tourist accommodation establishments

- Regulation EC 692/2011
Validity

To what extent does the available mobile positioning data serve as valid indicators for real world facts to be covered in the relevant statistics sectors?

Looking at the official definitions

Main issue with definition of ‘usual environment’

Many of the listed differences are minor and their effect on the output is likely to be negligible.
Accuracy: Coverage

Problems inherent in mobile data

Over- and under-coverage of aspects like:

- visitors who do not use mobile phones
- some visitors use more than one mobile device
- visitors not actually entering/exiting the country of reference (not crossing the border)
- ...

Careful evaluation and calibration using other sources required
Accuracy: Measurement and Processing

Missing values, incorrect formats, duplicated data

Mostly due to extraction from MNO databases, but are probably not more problematic as other data sources

Processing massive amount of data - solved with hardware scaling

Processing usual environment - Positium’s anchor point model is 95% correct on LAU2 level
Comparability

Comparability over time:

- Depends on changes in data quality
- Depends on changes in the telecommunication market (e.g. cost of calls/SMS, emerging of new MNOs, merging of MNOs)
- Roaming to be free in Europe as of 2016

Comparability between other countries
Synergies with Other Fields of Statistics

Possibilities for joint processing for:

Balance of payments
Transportation statistics
Population statistics
Methodology: Conclusions

Quality assessment relies heavily on existing external information.

Indicators do not comply to requirements of the regulation (692/2011) fully.

Longitudinal data

Coverage issues most important.

No easy estimation as no reliable reference data.
Questions?
TASK 3B: COHERENCE
Coherence of Mobile Positioning Data in Tourism Statistics

- Objectives and key questions
- Mobile positioning and reference data
- Framework for testing coherence
- Analysis
- Conclusions
Objectives and Key Questions

• Feasibility of mobile positioning data in the context of official tourism statistics

• In-depth testing of tourism statistics compiled based on mobile positioning data

• Analysis of coherence compared to reference statistics

• Key questions to be addressed:
  o Domain coverage: inbound, outbound and domestics tourism?
  o Tourism breakdowns: same-day and overnight trips?
  o Coherence to existing statistical indicators? Reasons for deviations?
Mobile Positioning Data

- Mobile positioning reference data
- Based data from Estonian MNOs between Jan 2008 and July 2013
- Up to 2010 based on data from a single MNO
- From 2011 based on data from two MNOs
- MNO data were processed into tourism statistics
- Inbound, outbound and domestic flows
- Origins and destinations
- Number of same-day and overnight trips
- Length of trips
## Reference Data - Tourism Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mobile positioning data</th>
<th>Supply statistics (accommodation)</th>
<th>Demand statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target population</strong></td>
<td><strong>Outbound / domestic:</strong> Population of the reference country</td>
<td>Accommodation establishments (all or above threshold)</td>
<td>Population over 15 years</td>
</tr>
<tr>
<td></td>
<td><strong>Inbound:</strong> non-resident tourists</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Frame</strong></td>
<td>Data of mobile phone subscribers</td>
<td>Business / tourism register.</td>
<td>Population register or area frame</td>
</tr>
<tr>
<td><strong>Source data</strong></td>
<td>Administrative</td>
<td>Enterprise survey</td>
<td>Survey of individuals</td>
</tr>
<tr>
<td><strong>Sampling design</strong></td>
<td>Census / sample</td>
<td>Census / sample</td>
<td>Sample</td>
</tr>
<tr>
<td><strong>Time units available</strong></td>
<td>Day / week / month etc.</td>
<td>Month</td>
<td>Quarter / Year</td>
</tr>
<tr>
<td><strong>Regional areas</strong></td>
<td>Any customized area</td>
<td>NUTS 2</td>
<td>Country</td>
</tr>
<tr>
<td><strong>Nationality breakdown</strong></td>
<td>Possible</td>
<td>Possible</td>
<td>Only residents</td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td>1-2 weeks</td>
<td>5-8 weeks</td>
<td>7-8 weeks</td>
</tr>
<tr>
<td><strong>Legal basis</strong></td>
<td>None</td>
<td>Regulation 692/2011</td>
<td>Regulation 692/2011</td>
</tr>
</tbody>
</table>
## Reference Data - Related Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mobile positioning data</th>
<th>Border interview</th>
<th>Passenger statistics</th>
<th>Border control</th>
</tr>
</thead>
</table>
| **Target population**   | **Outbound / domestic:** Population of the reference country  
**Inbound:** non-resident tourists | Inbound visitors to the reference country. | All passengers by transport mode (air / sea / train etc.) | All passengers passing through border control. |
| **Frame**               | Data of mobile phone subscribers | Main border crossing locations | Register of transport authorities. | Register of border authority. |
| **Source data**         | Administrative | Survey of individuals | Administrative | Administrative |
| **Sampling design**     | Census / sample | Sample | Census | Census |
| **Time units available**| Day / week / month etc. | Month / quarter | Day / week / month | Day / week / month |
| **Regional areas**      | Any customized area | Country | Country | Country |
| **Nationality breakdown** | Possible | Possible | Not possible | Possible |
| **Timeliness**          | 1-2 weeks | 10-20 weeks | 1-2 weeks | 1-2 weeks |
Method and Indicators

• Mirror statistics method is used for comparing mobile positioning to reference data

• Coverage indicator (CC2. Asymmetry for mirror flows statistics - coefficient)
  → Measures the absolute difference between two mirror flows

• Consistency indicator (Pearsons r)
  → Measures the linear correlation between the mirror flows where value 1 indicates perfect correlation
## Framework for Coherence Tests

<table>
<thead>
<tr>
<th>Tourism Domain</th>
<th>Mobile Positioning Data</th>
<th>Reference (Mirror) Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined inbound and outbound tourism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total trips</td>
<td>Inbound+outbound</td>
<td>Ferry passengers</td>
</tr>
<tr>
<td>Inbound tourism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total trips</td>
<td>Total trips</td>
<td>Demand Statistics (FI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Border Control (EE)</td>
</tr>
<tr>
<td>Overnight trips</td>
<td>Overnight trips</td>
<td>Demand Statistics (FI)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supply Statistics (EE)</td>
</tr>
<tr>
<td>Same-day trips</td>
<td>Same-day trips</td>
<td>Demand Statistics (FI)</td>
</tr>
<tr>
<td>Nights spent on overnight trips</td>
<td>Overnight trips</td>
<td>Supply Statistics (EE)</td>
</tr>
<tr>
<td>Outbound tourism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total trips</td>
<td>Total trips</td>
<td>Demand Statistics (EE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Border Interview (FI)</td>
</tr>
<tr>
<td>Overnight trips</td>
<td>Overnight trips</td>
<td>Demand Statistics (EE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supply Statistics (EU)</td>
</tr>
<tr>
<td>Same-day trips</td>
<td>Same-day trips</td>
<td>Not available (begins 2014)</td>
</tr>
<tr>
<td>Domestic tourism</td>
<td></td>
<td>Demand A</td>
</tr>
<tr>
<td>Total trips</td>
<td>Total trips</td>
<td>Demand Statistics (EE)</td>
</tr>
<tr>
<td>Overnight trips</td>
<td>Overnight trips</td>
<td>Demand Statistics (EE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supply Statistics</td>
</tr>
<tr>
<td>Same-day trips</td>
<td>Same-day trips</td>
<td>Not available (begins 2018)</td>
</tr>
</tbody>
</table>
Coverage indicator is positive for Accommodation statistics they include only paid accommodation. Total trips in Demand statistics have better coherence than overnight and same-day trips.
Mobile positioning data provides very good consistency overall and broken down into countries. There is slight over-coverage as more trips are registered.
Example: Very Good Coherence

Inbound + Outbound: Ferry passengers, FI <-> EE

Most passengers travel by ferry between Finland and Estonia. Consistency is excellent. Number trips is less in mobile positioning → other nationalities (than FI and EE) and transit passengers are on the ferry.
Mobile positioning data contains many such outbound trips that do not qualify as tourism trips in the Demand Survey either due to frequency, purpose or duration of the trip.
Border control registers all trips, regardless of purpose. Very short trips close to the border may be seriously underestimated in mobile positioning data because these tourists might not use the roaming services of MNOs in the inbound country.
Example: Domestic Trips Outside Usual Environment

Using LAU-2 for defining usual environment
Using LAU-1 for defining usual environment
Coherence: Strength/Weaknesses

Main strengths of mobile positioning data
1. Excellent consistency over time for the number of trips and nights spent
2. Superior coverage for overnight trips when compared to Supply Statistics: covers also trips in non-rented or non-registered accommodation
3. Possibility to produce breakdowns based upon region and nationality
4. Possibility to apply rules for usual environment
5. Many short trips are excluded from mobile positioning data

Main weaknesses of mobile positioning data
1. No additional data about the trip (purpose, expenditure, accommodation, means of transport)
2. Potential problems in the method for accurate breakdown of trips into same-day and overnight trips
3. Over-coverage issues related to usual environment (purpose, duration or frequency of trip)
4. Under-coverage issues based on mobile phone use: some tourists don’t use their phone abroad
Coherence: Conclusions

• Mobile positioning data alone cannot fulfill the requirements of the regulation (692/2011)

• Mobile positioning data provides good estimates for the number of trips, nights spent and destination

• Mobile positioning data doesn’t produce information on purpose of trip, type of accommodation or expenditure → need for traditional surveys
Questions?
TASK 4: OPPORTUNITIES & BENEFITS
Assessment dimensions

Mobile positioning data compared to traditional sources of tourism statistics to be assessed regarding ...

1. Quality  
   (basis: Reg. (EU) 692/2011)

2. Cost  
   (basis: Generic Statistical Business Process Model (GSBPM))

3. Access  
   (new indicators outside the Reg. (EU) 692/2011)

4. Synergies  
   (with statistical domains outside tourism)

5. Transmission
## Findings: Quality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>No complete coverage of any sector relevant for tourism statistics → No replacement of traditional sources</td>
<td>-</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Full integration and automatisation → Much quicker than traditional sources</td>
<td>+</td>
</tr>
<tr>
<td>Validity</td>
<td>No specific advantages/disadvantages</td>
<td>±</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Advantages over traditional sources (Smaller sampling error, no memory gaps) - need to re-define ‘usual environment’</td>
<td>+</td>
</tr>
<tr>
<td>Consistency</td>
<td>High grade of consistency in variation compared to traditional sources.</td>
<td>+</td>
</tr>
<tr>
<td>Resolution</td>
<td>Finer granulation of space and time → new possibilities (again, need to re-define ‘usual environment’)</td>
<td>+</td>
</tr>
</tbody>
</table>
Quality Implications

At present, mobile positioning data cannot replace traditional sources of tourism statistics but could deliver additional information...

1. Quick indicators
   (variation of key indicators of tourism statistics much faster than today)

2. Finer spatial and timely resolution than possible today

3. Source of calibration for traditional sources
   (necessary to quantify the bias between sources)
Findings: Cost

Example:
Country with a population of about 16 million, 3 MNOs (10, 5, 1 million subscribers), 15 days latency:

A: data processing carried out by MNOs

B: data processing carried out by NSI

Figures in ,000 EUR
## Cost of Processing in Comparison

### Mobile positioning data vs Current methods in Tourism Statistics

<table>
<thead>
<tr>
<th></th>
<th>Case example</th>
<th>Estonia</th>
<th>Finland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual man-days</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td>Demand-side</td>
<td>Demand-side</td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Workload</strong></td>
<td>264</td>
<td>600</td>
<td>690*</td>
</tr>
</tbody>
</table>

### Additional Information

<table>
<thead>
<tr>
<th></th>
<th>Case example</th>
<th>Estonia</th>
<th>Finland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 million inhabitants</td>
<td>1.3 million inhabitants</td>
<td>5.4 million inhabitants</td>
</tr>
<tr>
<td><strong>Data scope</strong></td>
<td>Inbound, outbound, domestic</td>
<td>Outbound, domestic</td>
<td>Inbound, domestic</td>
</tr>
<tr>
<td><strong>Update period</strong></td>
<td>Monthly</td>
<td>Quarterly</td>
<td>Monthly</td>
</tr>
<tr>
<td><strong>Collection method</strong></td>
<td>Automated data collection</td>
<td>Survey</td>
<td>Data request</td>
</tr>
</tbody>
</table>

*½ of the total workload as shared with the Consumer Survey
Cost Implications

1. Quite high implication cost - quite low annual running cost once the system is implemented

2. Less overall cost if processing of the data happens within the NSI rather than within the MNOs

3. Annual workload with mobile positioning data can be considerably less compared to the current methods of gathering tourism statistics

4. Potential for efficiency increase by mixed-mode data collection (demand survey augmented with mobile positioning data)
Findings: Access

New indicators based on mobile positioning data WITHIN official tourism statistics ...

1. Quick indicators
2. Source of calibration for traditional sources
3. Accommodation below the threshold/non-paid
4. Demand-side: below 15 years; outside ‘leisure’
5. Longitudinal analysis
Findings: Access

New indicators based on mobile positioning data OUTSIDE official tourism statistics ...

1. Space: Visitors of a customised location/region
2. Time: Visitors during a customised period
3. Space and Time → Event monitoring
4. Accommodation outside official tourism statistics
5. Longitudinal data (repeating/first time visitors)
6. Segment data (structural data about visitors)
Findings: Synergies

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

1. The Balance of Payments Statistics, specifically the travel item

2. Transport Statistics for additional calibration of origin-destination matrices

3. Migration and commuting statistics

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

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

Diagram:
- Mobile Network Operator (MNO)
- Data Extraction
  - Frame Formation
  - Data Compilation
  - Estimation
  - Combining
- Reference Data
- National Statistical Institutes (NSI)
- European Commission (EUROSTAT)
  - European Virtual Tourism Observatory (VTO)
- Deployment to Public
1. Fully automated transmission of the data plays a vital role in reducing manual work with the data and as a consequence getting gains in timeliness. Source of calibration for traditional sources.

2. The manual work has to be carried out in the validation, quality assessment and interpretation of the results.

3. Concerning the transmission standards, it is obvious that those should be set on the country-level, and follow as much as possible any kind of international standards.
Questions?
CONCLUSION AND IMPLICATIONS
Summary of Main Findings

Mobile positioning data as a source for tourism statistics ...

1. Limited data access possibilities due to regulatory limitations. Differences between EU countries.

2. Need for a central legal and methodological framework

3. Longitudinal data is a MUST

4. At present, rather a supplement than a replacement source

5. However, many possible benefits, e.g. in quality, cost, access to new indicators and synergies with other domains
## Strengths and weaknesses of mobile positioning data

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good consistency</td>
<td>Access/continuity of access</td>
</tr>
<tr>
<td>Superior coverage compared to supply statistics</td>
<td>No information on the purpose, expenditure, means of transport</td>
</tr>
<tr>
<td>Breakdowns by region and nationality</td>
<td>Bias between some classifications (e.g. same-day/overnight)</td>
</tr>
<tr>
<td>Various quantitative criteria for definitions</td>
<td>Possible misclassification of actual tourism events</td>
</tr>
<tr>
<td>Improved timeliness</td>
<td>Over- and under-coverage issues concerning the phone usage patterns</td>
</tr>
<tr>
<td>Automation level of statistical production</td>
<td>Difficulty to assess the accuracy of data as mobile phone usage on travel is unknown</td>
</tr>
<tr>
<td>Possible positive cost effects</td>
<td></td>
</tr>
<tr>
<td>Pan-European travel network statistics</td>
<td></td>
</tr>
</tbody>
</table>
Balance of travels between two countries

- Direct travel
- Travel through transit countries
Balance of travels between many countries provides pan-European travel network
### Implications for the Statistical Community

#### EU Level
- ESS will benefit from the knowledge and case studies contained in this study
- Eurostat to follow the developments
- Eurostat to act as a scout for application

#### NSI level
- Verify their country’s position on the continuum from ‘very liberal’ to ‘very restricted’ in terms of privacy protection
- Determine the specifics for the situation to be monitored
Final Words

1. There are many potential benefits in the use of mobile positioning data for Eurostat, NSIs, MNOs and users of tourism statistics.

2. There are many barriers that need to be overcome.

→ There is a lot of work ahead!

→ This study can act as a common starting point!
Thank You!

Acknowledging:

Respondents of the Survey & Interviews

MNOs providing the data (TeliaSonera / Elisa)

Czech Statistical Office

Eurostat

Consortium members

Project website: mobfs.positium.ee