





coverage optimisation, marketing analysis, data retention laws, etc. An average person makes multiple calls and sends text messages every day and if possible, uses a data connection. Every day, millions of locations are stored. That means a small country generates billions of unique locations every month. The volume and velocity itself is enough to term this a big data source.

The data is very sensitive and can be considered subscribers' private data. However, the aggregated data about locations and movement patterns with various time, spatial, demographical, and other breakdowns represents a valuable resource for statistical indicators in several application domains. If processed according to privacy standards and data protection legislation, the confidentiality of the subscribers is preserved.

The data describes human behaviour historically over longer time periods. The insights are numerous, giving us an understanding about where people are, where they come from and how they move. The opportunity: identifying repetitive patterns in location activity over time enables to build a clear picture of the visitor profile of the location itself.

There are two ways of looking at mobile positioning data:

1. Segment insights. Knowing where people are over time.
2. Place insights. Knowing who and how many people are staying at or passing through a destination over time.

It means there are several domains where mobile positioning data can be used as a source of statistical indicators: tourism, population, transportation & commuting, urban & regional planning, geomarketing, public safety & security, international travel and academic research.

Information in this detail has been previously hard to obtain. However, with mobile positioning data it is now possible to leapfrog other legacy market research methods and measure direct behaviour based on existing data.

While big data makes its crusade, non-official statistics begin to appear and used widely. National statistical institutes (NSI) need to deeply consider their role in the big data era. NSIs fulfil a vital role in society as an impartial anchor. As researchers from Statistics Netherlands point out (Struijs & Daas, 2014), it is acceptable that others produce information based on big data, if this is validated by NSIs.

### **Eurostat feasibility study on the use of mobile positioning data for tourism statistics**

For data collection systems in the field of tourism the need for an update has been long underlined by both users and producers of the data. This has led Eurostat to invest in developing alternative sources of data.

Common sources of data in tourism include surveys filled in by accommodation providers (reporting on their guests) and surveys filled in by households and individuals (reporting on trips made during a recent reference period). Eurostat has recognized that the European Statistical Society business environment has changed, however. New data is required, collection needs to be simplified and the ensuing burden reduced. The European Statistical Society now also requires that data sets are joined and existing/administrative data is re-used. With this in mind, Eurostat needed to play a role in pursuing more efficient ways of collecting data, also in the field of tourism statistics.

Positium and the University of Tartu were part of a consortium of expert organisations for an assignment by Eurostat to determine the feasibility of mobile positioning data for tourism statistics (Eurostat, 2014). It aimed at exploring the possibilities – and limits – of using positioning data stored by mobile network providers for measuring tourism flows.

The project analysed barriers to access data. The methodology was described in depth. In addition, correlation and coverage analyses (vis-à-vis traditional statistics) showed the potential of mobile

positioning data. Positium and the consortium produced a thorough manual that lays out the steps necessary to access and apply mobile positioning data for producing tourism statistics. The work is a starting point for a unified central framework in Europe to use mobile positioning data in tourism statistics. The feasibility study discusses strengths and weaknesses of using mobile positioning data in official statistics. Main weaknesses of mobile positioning data are access to the data, methodological differences from traditional statistics and over- and under-coverage issues concerning phone usage patterns.<sup>1</sup>

There are three major benefits of mobile positioning data versus traditional methods:

1. The data is already there, passively collected through normal use of the mobile device – not requiring any apps to be installed on the phone and is gathered from any type of mobile device. Continuous long-term data allows us to analyse a longer term and detect longitudinal patterns.
2. The data has national coverage, but is accurate enough (in cities within one city block radius) to provide local insights.
3. Mobile phone usage is wide spread so almost everyone are sampled in the data, including visitors and those passing through.

The advantage of passive mobile positioning is that it creates the opportunity to analyse visits of a large part of the population over a long period of time. Because it is a very large but high fidelity database, we can in turn segment this data in many dimensions.

The following dimensions can be used for segmenting:

- Geography (Where?)
- Timing (When?)
- Origin (Where from?)
- Duration (How long?)
- Frequency (How often?)

The result is that we can look at the distribution of trips by the timing and geography of the visits. We can also include information on the frequency or duration of trips into the segmentation process. The duration of trips can, in turn, be calculated in terms of one trip, month or year.

The Eurostat feasibility study on mobile positioning for tourism statistics summarizes major indicators in tourism that are available in Table 1.

Table 1. Major indicators in tourism available using mobile positioning data (Eurostat, 2004).

Field of application	Possible indicators	Remarks
Detailed <i>spatial</i> statistics	Number of visitors at a customised location	
	Share of different nationalities	
	Duration of stay	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data
Detailed <i>temporal</i> statistics	Number of visitors during a customised period	Such as defining day visits/multi-day visits
	Share of different nationalities	
	Duration of stay	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data

<sup>1</sup> See the study (Eurostat, 2014) for a thorough discussion of the data source.

Detailed <i>spatiotemporal</i> statistics (e.g. event monitoring)	Number of visitors at a customised location during a customised period	Such as event monitoring
	Share of different nationalities	
	Duration of stay	Requires longitudinal data
	Repeat/first-time visits	Requires longitudinal data
	Visitor segments	Requires subscribers data
Segment data	Residence, sex, age, family status, etc.	Requires subscriber information

### Applications – Spatio-temporal segmentation for event discovery and large event analysis

Based on the indicators outlined in the previous chapter it is possible to offer tourism statistics at the general administrative level, but also for specific temporal and spatial phenomenon with their unique setting, people and management system – events (Getz, 2005). Getz argues that events are an important motivator of tourism and figure prominently in the development and marketing plans of any tourism destination. As it has been very difficult to obtain valuable event visitor statistics with traditional methods, we will outline two event application examples from Estonia that are based on mobile positioning data and could fill the gap for this kind of data collection. We also propose an application for retroactively analysing the World Cup of 2014 in Brazil and monitoring 2016 Olympic Games in Rio de Janeiro.

#### Estonian Song and Dance Festival

The Estonian Song Festival is one of the largest amateur choral events in the world, a masterpiece of the oral and intangible heritage of humanity. The festival is held every five years in July on the Tallinn Song Festival Grounds.

The organisers estimate the number of visitors to 30 000 singers and 80 000 visitors. However there is very little idea where the visitors came from and which days they actually visited the festival.

The organisers of the Song Festival are using following type of mobile positioning data of foreign visitors for their planning and marketing purposes:

- The number of foreign visitors during the festival
- Number of first time visitors and repeat visitors by country of origin – top 10 countries
- Distribution of guests by duration of stay, broken down by
  - One-day visitor top 10 countries of origin; Multiday visitors top 10 countries of origin
  - One-day visits, 2-3, 4-6, 7-14 and longer stays
- Other destinations of visitors in Estonia
- Comparison data on previous Song Festivals

#### Rally Estonia

Having an international rally event attracts a lot of people from the region and across borders and boosts the local economy. But admission to the stages is mostly open. The question then becomes, how effective are the events in attracting tourists and where do visitors come from?

The Rally Estonia is a rally event held in South Estonia. It is the biggest annually organised rally event in the Baltic countries and part of the 2014 European Rally Championship season.

Rally Estonia needs information to improve on its marketing activities from year to year and secure new long-term contracts with sponsors.

Mobile positioning data helps them to:

1. Pinpoint the number of visitors for each of the stages of the rally event

- Determine the origin country (when foreign visitor) and municipality (when domestic visitor) for visitors for the entire rally.

Based on this data Rally Estonian organisers were able to present valuable information for their sponsors and measure the overall expenditure created by Rally Estonia in the region where it took place.

### World Cup 2014 and the 2016 Olympic Games in Rio de Janeiro

The Instituto Brasileiro de Geografia e Estatística (IBGE) is, as the national statistical institute, the main provider of data and information in Brazil. With this in mind, IBGE is looking to conduct a project to pilot an innovative methodology for the production of tourism statistics to determine the tourism flows in Brazil at the time of the World Cup (June-July 2014) based on mobile positioning data. Currently, data on World Cup visitors is severely lacking.

IBGE is foremost interested in the following from the pilot with mobile positioning data:

- Volume of tourists in each World Cup city daily
- Percentage co-visitation and the order of travelling between multiple World Cup cities
- Breakdown by country of origin
- Testing a new method of data collection and comparing the results to traditional sources (hotel reservations and entrants to Brazil)

To prepare for analysis of mobile positioning data we are proposing a prototype solution to act as a user interface for the statistics. This prototype can similarly be applied to monitor the forthcoming 2016 Olympic Games in Rio de Janeiro (see Figure 2).

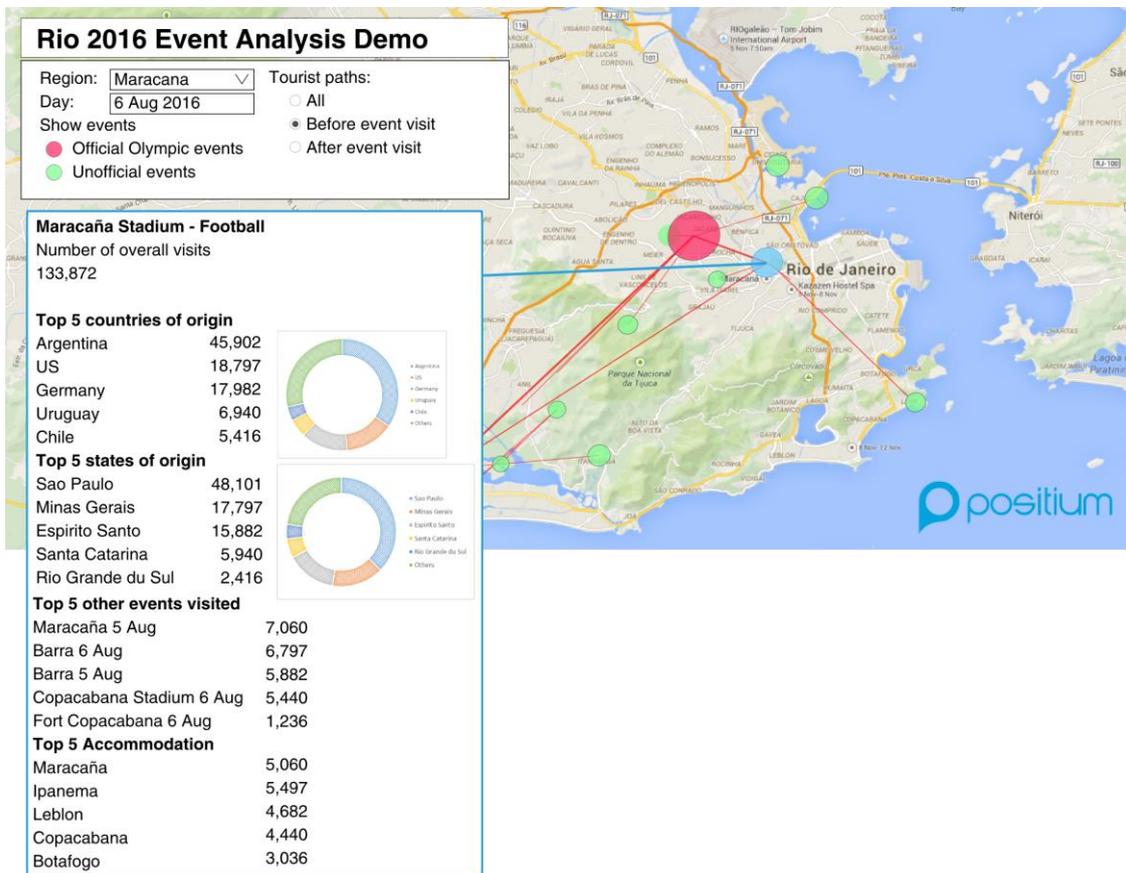


Figure 2. Prototype of a potential application for event analysis with dummy data (Positium).



This proposed application meets two overarching objectives:

1. Giving an overview of all mass-gatherings and events in the area in the given time period
2. Providing detailed statistics of any given event
  - Number of people at the place of event
  - Origin of the people – separating the origin of domestic and foreign tourists
  - Accommodation – where are the people at this event staying overnight in the city
  - Cross-visitation – other events in the city visited during the same period

Normally, it is hard to meet requirements for both a good overview and the detail necessary for correct analysis of the events. With this setup the demo application already covers a wide spectrum of feature requests from potential users. In addition, a separate XLS file can be generated for each of the events during the custom time period, further strengthening the possibility to run detailed analysis.

## Conclusions

This paper introduced a source of data for tourism statistics. The main argument is that this source of data is granular enough to be segmented into time and space categories of choice. Questions like “How many people visited the city of Manaus on the day of the World Cup match? Where did they come from?” can now be answered from a tourism statistics database. We also presented a mock-up of an event analysis application that enables to provide a general overview and also investigate the specific statistics of a time and place.

As with any source of data, mobile positioning data is not an oracle that answers all the questions, but it does answer some question in much more detail. Mobile positioning data can be seen as a necessary quantitative complement to qualitative surveys. It provides significant insights on its own, and this quality is extended in a mixed mode with surveys.

National statistical institutes are figuring out that they need to stay ahead of the curve with new ways of gathering statistics. When successful, NSIs will remain the go-to place for spatio-temporally segmented tourism statistics. Today, data is being produced everywhere and statistical institutes are doing well to embrace Big Data to stay the forerunner in producing statistical insights.

## References

- Eurostat (2014). Feasibility study on the use of mobile positioning data for tourism statistics. [http://epp.eurostat.ec.europa.eu/portal/page/portal/tourism/methodology/projects\\_and\\_studies](http://epp.eurostat.ec.europa.eu/portal/page/portal/tourism/methodology/projects_and_studies).
- Getz, D. (2005). Event Management and Event Tourism. New York-Sydney-Tokyo: Cognizant Communication Corporation 439.
- ITU World Telecommunication/ICT Indicators database (2014). <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.
- Struijs, P. & Daas, P. (2014). Quality Approaches to Big Data in Official Statistics. UNSD International Conference on Big Data for Official Statistics. <http://unstats.un.org/unsd/trade/events/2014/Beijing/documents/other/Statistics%20Netherlands%20-%20Quality%20Approaches%20to%20Big%20Data.pdf>.