12th World Telecommunication/ICT Indicators Symposium (WTIS-14)

Tbilisi, Georgia, 24-26 November 2014

Background document	**Document INF/2-E** **18 November 2014** **English**

SOURCE:	Statistics Netherlands
TITLE:	Innovation of tourism statistics through the use of new big data sources

12th World Telecommunication/ICT Indicators Symposium (WTIS-14)

Tbilisi, Georgia, 24-26 November 2014

Background document	**Document INF/2-E** **18 November 2014** **English**

SOURCE:	Statistics Netherlands
TITLE:	Innovation of tourism statistics through the use of new big data sources

Innovation of tourism statistics through the use of new big data sources

Nico Heerschap, Shirley Ortega, Alex Priem en May Offermans, *Statistics Netherlands,* Netherlands

Email: nhrp@cbs.nl

1. Introduction

In the last decade, more and more data have been becoming available for research and development. There seems to be a move from data scarcity to data abundance ("data deluge"). This is primarily the result of increasing possibilities to digitize growing volumes of data at lower costs. Not only large administrative data sources are becoming available; data are also being generated by sensors (e.g. cameras, electronic traffic loops and public transport cards), mobile telephony and radio frequency identification (RFID chips). Most of these data are collected to control operational processes and for management. However, the data are increasingly being used for purposes other than those for which they are collected, such as the analysis of consumer behaviour for marketing strategies, predicting trends, detecting fraud, and also to produce new, faster and more detailed statistics. The development of the internet is a second important factor in this respect, especially the transition from web 1.0 to 2.0 at the beginning of the century. Internet users are no longer passive recipients of content; they absorb information from the internet and in return produce their own new content which in turn is spread via the internet. This change from unilateral to multi-way communication has also been the breeding ground for the emergence of social media. This is, finally, enforced by the introduction of mobile internet supported by smartphones and other portable devices. All these activities leave behind some form of digital footprints that can be detected, measured and analysed. This kind of analysis falls under the term "big data". Although not clearly defined, big data relates to datasets which are so large and complex in structure that they cannot be processed and analysed with traditional technology, software and methodology [1, 2].

This paper presents three examples of innovative (big) data research for tourism statistics. They were carried out by Statistics Netherlands (SN) in the period 2012–2013. The first example is the use of the internet for the compilation of the population of units for the tourism accommodation statistics, which produce indicators on inbound and domestic tourism. The second example focuses on the use of log data registered by an app installed on the mobile phones of a group of respondents (smartphone measurements). And the third example consists of an analysis of mobile metadata, which are generated when a mobile phone communicates with a telecom provider, so-called Call Detail Records. In the latter case Vodafone and Mezuro supplied anonymous and aggregated data. The paper concludes with some general remarks about the effects of these new and innovative data sources for the production of tourism statistics.

2. Internet as a source for tourism statistics

The internet has become part of society's basic infrastructure. More and more activities are taking place through the internet, leaving behind digital footprints which can be detected and measured in real time. This also applies for tourism accommodations and their characteristics, such as hotels, campsites, youth hostels, bed and breakfasts etc. This information can be collected from the internet by so-called robots or crawlers. Internet robots are small software programs which trawl the internet gathering the desired information (web page content). This can be done in one of two ways. Either with a *dedicated robot,* which gathers information from a specific website, such as Booking.com, Tripadvisor.com or other specific tourism websites where tourism accommodation providers post their information. Or with *general robots* that search or scrape the internet, or part of the internet, looking for web pages which contain designated keywords, such as "hotel", "accommodation" or "B&B"; these robots are not built for or limited to specific websites.

SN experimented with both methods to compile the population of units for the tourism accommodation statistics. The internet data collection focused on the smaller accommodations. The method was also applied to find the tourism accommodation units in the Caribbean Netherlands, where tourism accommodation data were incomplete. The Chamber of Commerce was not able to provide a usable set of population units.

The main conclusions of these pilots were:

- Technically it is not very difficult to do. The robots required, especially the dedicated robots, can be built and tested within a couple of days. This also applies if robots have to be updated if websites change their layout or technical structure.

- In addition to names and addresses of tourism accommodations, these methods can also collect various other characteristics: number of rooms, prices, tourist tax, available facilities and guest review scores, but also job vacancies and Chamber of Commerce registration numbers (see Table 1). The latter can be used to link the collected internet data to other relevant statistics like turnover, investment and innovation. The internet information can also help to improve the classification of accommodations.

Table 1. Example of data collected on tourism accommodations by a dedicated internet robot (fictitious data).

Name	TypeAcc	#Rooms	#Beds	Price p/n	City	Street	Num	Tel	Star rating	Swim. pool	Free wifi	Job vacancy	Website
Hotel Amsterdam	Hotel	174	400	$125-200	Kralendijk	J. W. Boulevard	18	(+599) 720 123	4	y	y	2	www. amsterdam.com
Henk's Beach Resort	Hotel Resort	102	210	$201-324	Hato	Kaya Apeldoorn	11	(+599) 720 567	4	y	y	8	www.henk.com
Dive Hotel	Hotel	38	76	$104-117	Kralendijk	Kaya Nice	40	(+599) 722 901	2	n	y	0	www.dive.com
Casino Resort	Hotel Resort	56	112	$120-200	Kralendijk	J. W. Boulevard	25	(+599) 723 098	3	y	y	3	www.club.com
Club Apartments	Appartment	40	80	$37 - 115	Kralendijk	Kaya Gold	104	(+599) 798 654	5	y	n	0	www.villas.com
Crown Villas	Villa	24	48	$231-403	Sabadeco	Lagoen road	80	(+599) 745 210	3	y	y	0	www.crown.com
Belair Apartments	Apartement	12	24	$235	Sorobon	Lighthouse	11	(+599) 745 123	2	n	y	1	www.belair.com
Hotel Rest	Hotel	10	22	$104-117	Kralendijk	Hill side	2	(+599) 723 456	2	y	n	3	www.rest.com
Eco Villa	Villa	7	16	$85-215	Belnem	Kaya Mooi	17	(+599) 745 123	4	y	n	0	www.eco.nl
Prince guesthouse	B&B	5	10	$90-125	Kralendijk	Kaya Prince	5	(+599) 723 456	1	n	y	n.a.	www.prince.com
Casa White	Studio	3	8	$200-250	Kralendijk	Kaya Red	8	(+599) 723 789	3	y	y	1	www.white.com
Villa Rose	Villa	2	4	$650-700	Hato	Flower hill	1	(+599) 745 012	5	y	n	0	www.rose.nl
Residence Yun	Appartement	2	4	$65-115	Sabadeco	J. W. Boulevard	19	(+599) 745 098	4	y	y	0	www.yun.com
Cottage High	Appartement	1	3	$100-150	Kralendijk	Kaya Linda	3	(+599) 745 876	1	n	n	n.a.	www.hill.nl

- Dedicated robots which collect data from specific websites result in better information than general robots which trawl the "entire" Internet. The advantage of dedicated robots is that we know beforehand what kind of information is available on the selected website and how the website has structured this information. The crucial step in this respect is then to find the group of websites which covers the whole domain as best as possible. Generally speaking, these include aggregation sites, such as Booking.com or Expedia.com, or specific tourism sites for, for example, campsites, holiday homes, hotel chains or bed and breakfasts. Although general robots cover the "entire" Internet, in this case the challenge is to determine whether a collected web page contains information on an actual tourism accommodation, and not, for example a web page of a hotel construction company, investment company or a general tourism site, which also contains words like "hotel", "accommodation" etc. The collected information is often incomplete and much more unstructured.

Table 2. Example of information problems of data collected on tourism accommodations by an internet robot (fictitious data).

Name	Typacc	#Rooms	#Beds	Price p/n	City	Street	Num	Tel
2 bedroom house in Hato	Cottage	2	4	$50-100	Hato	Kaya J. Willem Boulevard		
3 bedroom oceanfront penthouse	Apartment	3		$210-260	Kralendijk			(+599) 723 345
3 bedroom villa - ocean views	Villa	3	7		Sorobon	J. Willem Boulevard		
Affordable charming house	Villa			$87	Kralendijk	Near the Lighthouse		
House in Sabadeco	Apartment	2	4	$95-135	Sabadeco			
Spacious condo in Nawati	Studio	1	2		Nawati	J. W. Boulevard	3	Contact Richard
Beautiful renovated apartments	Apartment			$85	Kralendijk			(+599) 722 123

- The major challenge is not collecting the data from the internet, but the next stage: processing them, i.e. extracting the desired information, cleaning them up and converting them into useful information for the production of statistics. For example, a tourist accommodation often will post its information on more than one website. It often has its own website, but will also post on several aggregation sites. This means that after the collection of the data, a de-duplication process has to be implemented, e.g. based on names or addresses of the accommodations. This is not an easy task in practice, as building a good de-duplication program requires some effort. This step becomes more complex when names and addresses of accommodations are not available, are spelled differently or when the information is ambiguous or out-dated (see Table 2). Other challenges are standardisation

4

of names and addresses; comparability and interpretation of prices, as it is not always clear which elements (e.g. cleaning, tourism tax and activities) are included in the price; calculating the number of bed places, as usually only the number of rooms or units (e.g. buildings or cottages) is available; determining how up-to-date and valid the information is (e.g. the website may still exist while the accommodation has been closed); the reliability of guest review scores (these may be biased or even fake).

- It is important to ensure continuously that the whole population is observed, and that the robots do not miss any accommodation groups. In the case of tourism accommodations, this is probably only an issue in theory: if a tourism accommodation can no longer be found on the internet, it is highly debatable whether it is still operational. However, in the case of dedicated robots and thus selected websites, websites in this group may no longer be updated as a consequence of a fall in popularity compared to other and new websites. Lastly, statisticians should also be aware that a tourism accommodation may change the way it posts its information on the internet, for example by using social media or advertising on eBay.

- Internet data can be collected in real time and are more comprehensive than data collected by traditional surveys and from other sources. However, to assess the quality of internet data it is always wise to compare them with data from other available sources. Also, the fact that information often overlaps between websites is helpful (part of the de-duplication process).

So the general conclusion is that the collection of information on tourism accommodations or in fact any other tourism supply - restaurants and museums, for example - from the internet by robots has some potential, as more information than name and address only can be collected at much lower costs. Therefore, using internet information can reduce the survey burden, especially for small businesses. Data are available faster, new statistics can be produced, and operational efficiency can be improved. However, practice shows that methods for the processing, interpretation and visualization of the data are not easy to implement and take time to develop. Moreover, the quality of the internet data and the question of whether the whole population is observed also remain issues. Therefore, at present, SN sees data collection by internet robots more as a supporting data source alongside already available sources, i.e. the Chamber of Commerce register and the Business Register. Internet data can be used for information on smaller accommodations and businesses whose main activity lies outside the tourism domain, such as farms with campsites. They can also be used to check the quality of the current population of units of the tourism accommodation statistics. However, information from the internet is currently not seen as the sole source of information to compile the population of units for the tourism accommodation statistics.

3. Using mobile phone data for tourism statistics

Two other innovative projects at SN in the domain of tourism used mobile phone data, i.e.: 1) log data from smartphone measurements and 2) anonymous aggregate mobile phone metadata based on Call Detail Records. The former project was carried out in collaboration with the Delft University of Technology and the latter in collaboration with Vodafone (telecom provider) and Mezuro.

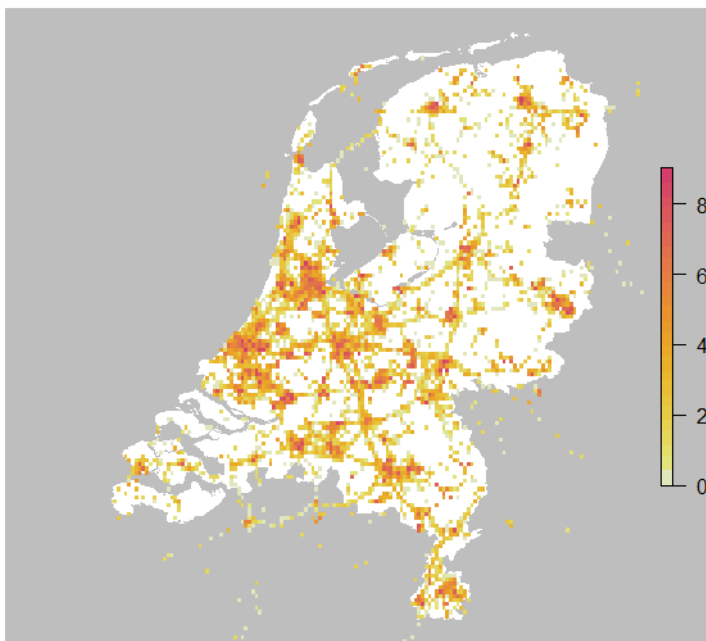## 3.1. Log data from smartphone measurements software

Log data from smartphone measurements are data collected directly from mobile phones of respondents through an *app installed on their phone or mobile device*. This app registers either everything or certain specific actions carried out on a mobile phone, including a time and location registration every five minutes. The latter two variables in particular are interesting for tourism statistics [3].

Besides the fact that movements of a person or family can be tracked quite accurately through time by GPS measurement, another advantage of this method is that specific questions - so-called pop-up questions - can be triggered on the basis of, for example, (the change in) time or location. This includes, questions about the purpose of the journey, mode of transport, price paid, type of accommodation, restaurant visits, satisfaction, activities etc. In more sophisticated apps, some of these variables can be derived on the basis of the location or registration of other data, such as mode of transport, journey duration and distance, mo-

bile payment and overnight stays. Another advantage of this method of data collection is that respondent background data are available: age, sex, income, region, composition of the family/group, making it possible to control the sample and weighting.

SN tested this method in two successive pilots in 2011 and in 2012. These pilots were not specifically conducted for tourism statistics but for mobility statistics and ICT use statistics (what do people do with their mobile phones?). The smartphone measurements were combined with two flanking surveys, one before the actual measurements and one at the end of the study. This was done to collect information about variables that cannot be registered on the basis of an installed app, to evaluate the project, and to compare this method of data collection with traditional surveys.

Figure 1. Visualisation (heat map) of the travel behaviour of a group of respondents based on log data from smartphone measurements, 10 October - 7 November 2012 [Bouwman et al].

Source: Statistics Netherlands and TU Delft.

Practice showed that the combination of information from traditional surveys and log data from smartphone measurements in particular produced a rich and valuable set of data with a lot of potential for research and statistics (see Figure 1, for example). Not only did this produce valuable data, the quality of the data was also improved considerably. A comparison between data registered by the app and the same data collected through the survey (self-report) at the end of the project showed a substantial number of wrong estimates in the self-reports (overestimation). A major challenge in this kind of research is the willingness of people to participate. Not only privacy concerns, but also technological issues and people's lack of knowledge about how to download applications played a role in this respect. The project dealt with privacy concerns by separating the data collection and analysis. The researchers did not have access to names and addresses of participants. In addition, participants could opt out anytime during the data collection.

In addition to the pilot described above, SN has also looked specifically at smartphone measurements software and tracking devices to support or replace questionnaires for the Mobility survey. This study was of a more technical nature, i.e.: to explore how accurate the mobility software was in the registration and determination of location, journeys and modes of transport [4].

3.2. Anonymous aggregated mobile phone metadata based on Call Detail Records[1]

In 2013 a pilot was conducted to see whether aggregated mobile phone metadata based on so-called Call Detail Records (CDR) could be used for tourism statistics. This study was carried out in collaboration with Vodafone and Mezuro. CDRs are generated by telecom providers through phone masts when mobile devices (e.g. mobile phones) make contact with the telecom provider or vice versa. For each contact (call, SMS or data session) these CDRs contain among other things an assigned identification number of the mobile device (not available for SN), the date and time of the call/SMS/data session, which kind of communication was used (voice, data or SMS), and location. The location is a reference to a radio cell in the operator's network, that is mapped to a geographical area. This is less accurate than a location measured with GPS. In addition, similar data were received from the telecom provider on the aggregated number of unique foreign mobile phone callers per month. In handling this kind of information, researchers must pay special attention to protecting and managing privacy by ensuring the anonymity of individuals. Therefore a special system was built and used to create aggregated, anonymised datasets.

The scope of the project was limited to inbound tourism specifically, because this group can be relatively well defined as a separate group. Specific data are also available from foreign mobile devices, i.e. so-called "roaming data" including the country code of the SIM card used. This makes it possible to gain an insight into times and places visited by groups of foreign tourists by country of residence, and to a lesser extent in the number of foreign visitors. To keep things simple, we assumed that every foreign device or unique mobile caller represented one inbound tourist, and that the country code of the SIM card was equal to the country of residence. It was also assumed that foreign mobile callers were randomly assigned to the nearest telephone mast and telecom provider[2].

To get a first indication of the quality of the data, the trend in the number of unique foreign mobile users of Vodafone NL Telecom services in the period April 2012 – March 2013 was compared to the trend in the number of foreign tourists visiting the Netherlands according to SN's survey-based tourism accommodation statistics for the same period. Similarly, a comparison was made between the trend in the aggregated counts obtained from the CDRs and the trend in the number of foreign visitors from the tourism accommodation statistics for the period January – March 2013 (see Figure 2). Subsequently, research was done for a selected set of specific cases concerning small tourist areas (e.g. the Dutch coast and specific events) and short time frames (e.g. calling patterns for a week and for national holidays).

The main conclusions of the pilot were:

- The *trends* in the data on the number of unique foreign mobile callers (blue line in Figure 2) and the data from the CDRs (green line in Figure 2) are comparable to the trends in the number of inbound tourists based on the tourism accommodation statistics (red line in Figure 2). In addition, predictions were made for the then still to be measured number of foreign tourists for the months April, May and June 2013 from the tourism accommodation statistics based on the mobile phone metadata. This was done by extrapolating the number of foreign tourists in March from the tourism accommodation statistics with the trends in the number of unique foreign mobile callers and in the CDR data. When the results (dotted lines in Figure 2) were later compared with the actual results from the tourism accommodation statistics, they only deviated by about 5 percent.

- The deviations between the *number* of foreign tourists according to the tourism accommodation statistics and the number based on the unique number of foreign mobile callers were quite large, for example 2.5 million and 3.5 million in the first quarter of 2013, respectively. However, this is not really surprising because, among other things:

    o the tourism accommodation statistics do not include foreign visitors in accommodations with fewer than 10 bed places and those staying with friends and relatives;
    o the number of unique mobile callers also includes day trips of residents from neighbouring countries, which account for many one-day visitors.
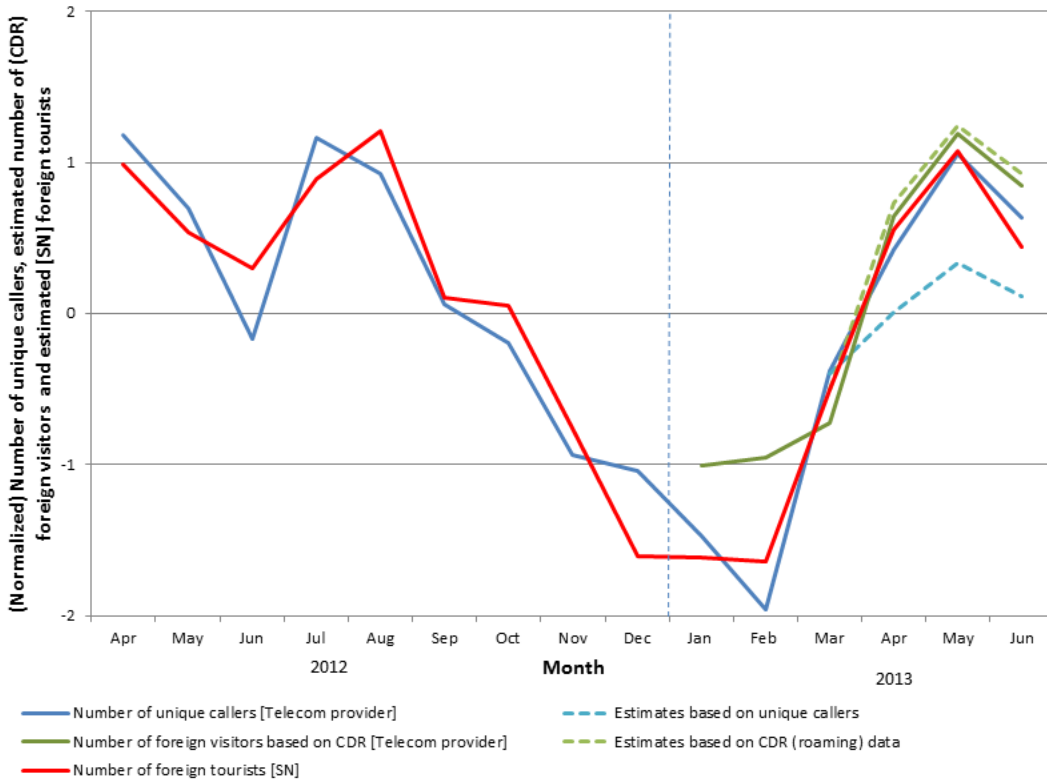
    Furthermore, it should also be taken into account that, besides Vodafone, there are other telecom providers in the Netherlands.

---

[1] Later this year a more elaborate paper will be published on this issue. See for other studies in this field for example Tiru and Ahas, 2013 [5].
[2] In a later stage of the study this proved not entirely to be true. This raises the question of representativeness.
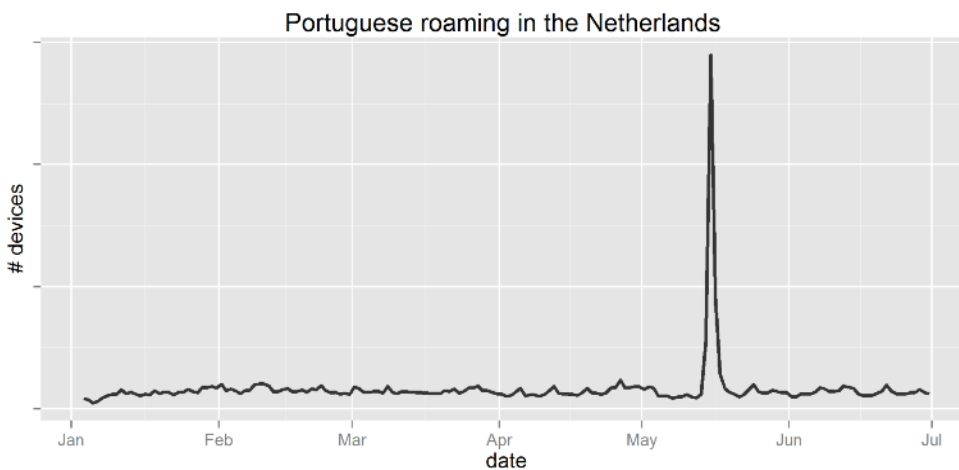
- The differences between the numbers of foreign tourists according to the tourism accommodation statistics and the numbers based on the CDR data are much larger. However, these numbers are not really comparable, as in the pilot a mobile device could only be followed for one day. Therefore, the next day the same mobile device was counted as a new device, resulting in double counting.

Figure 2. Comparisons between trends in inbound tourism based on SN Tourism accommodation statistics, unique foreign mobile callers, and CDR data[3].

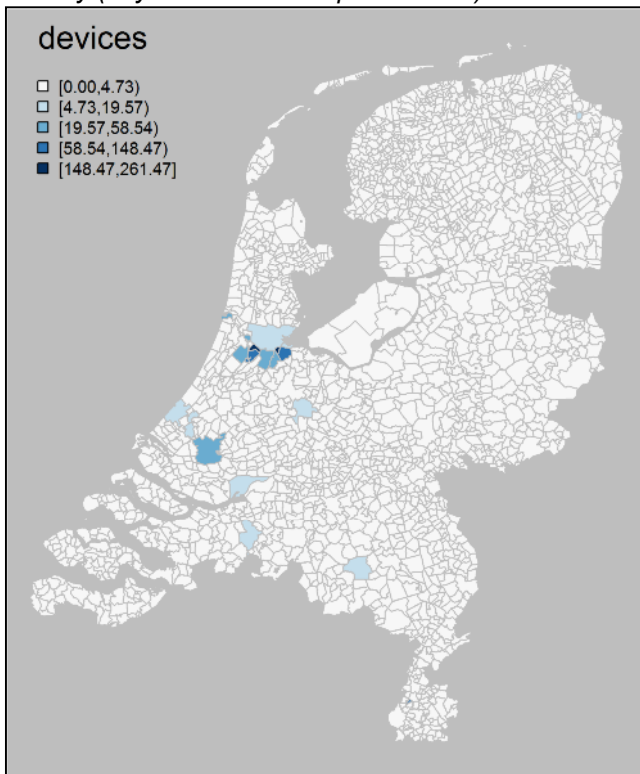*Source: Vodafone, Mezuro and Statistics Netherlands*

Figure 3. Call behaviour of Portuguese tourists (devices) in the Netherlands, January – July 2013.

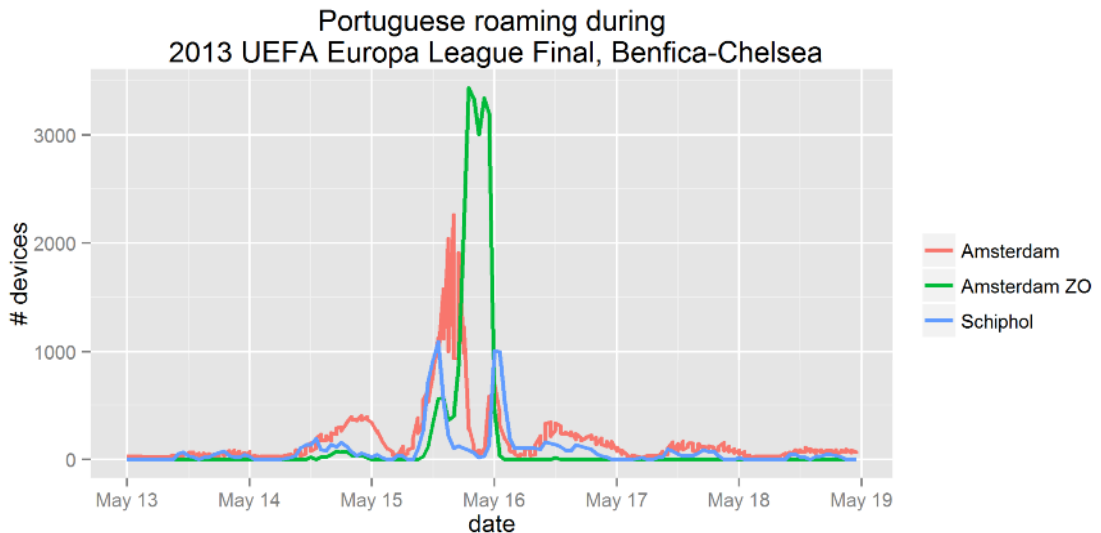*Source: Vodafone and Mezuro, compiled by Statistics Netherlands*

---

[3] To keep things simple and for the comparability of the trends in the data, all the time series have been standardised using their means and standard deviations. This ensures that all the time series have the same scale (same minimum and maximum and average zero).

8

*Figure 4. Call intensity of Portuguese tourists (devices) based on their **location** in the Netherlands on 15 May (day of the UEFA cup final 2013).*

Source: Vodafone and Mezuro, compiled by Statistics Netherlands

Figure 5. Calls by Portuguese tourists (devices) before, during and after the UEFA cup football final on 15 May 2013 (green line), Schiphol Amsterdam Airport (blue line) and Amsterdam in general (red line).
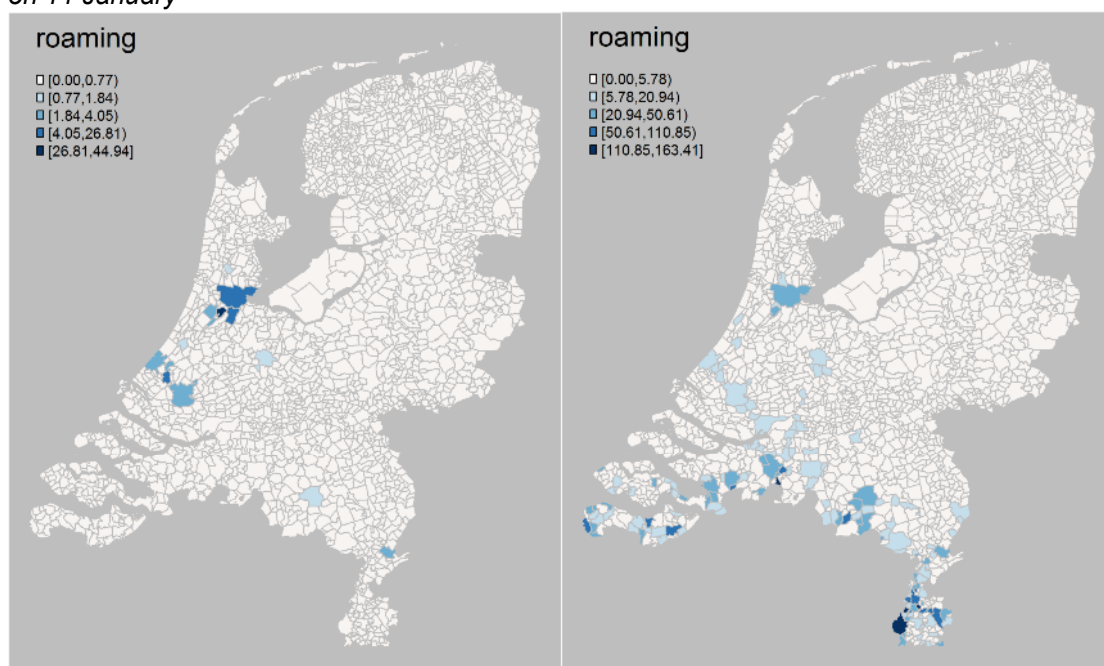

Source: Vodafone and Mezuro, compiled by Statistics Netherlands

- It proved quite easy to zoom in and present results on the patterns of call behaviour of foreign tourists on specific days (e.g. holidays like Easter) or in small areas (e.g. certain events or the Dutch coast). One specific example is given here: the UEFA cup football final between Benfica and Chelsea in Amsterdam on 15 May 2013 in the Arena Stadium in Amsterdam. Figure 3 clearly shows the peak in calls by Portuguese tourists (devices) on 15 May. This is also reflected in Figure 4, which shows

9

the intensity of calls by Portuguese tourists (devices) in the Netherlands on 15 May. The highest intensities are indeed in south Amsterdam, which is the location of the Arena stadium (Amsterdam ZO). Lastly, plotting calls by Portuguese tourists in Amsterdam and at Schiphol Amsterdam Airport, shows that apparently most Benfica fans arrived to watch the final and left the same day (see Figure 5).

- Figure 6 illustrates another example of the usability of mobile phone metadata. It shows the number of roaming events of Asian and Belgian tourists (= devices) in the Netherlands on 11 January from 12.00 – 16.00 hrs. The results show clearly that Asian tourists are concentrated in the big cities: The Hague, Rotterdam and Amsterdam, and Schiphol Amsterdam airport. As Belgium borders on the Netherlands to the south, roaming data of Belgian devices clearly show the expected border traffic. [6].

Figure 6. Roaming data for Asian (left) and Belgian (right) tourists (devices) from 12:00 – 16:00 hrs. on 11 January

Source: Vodafone and Mezuro, compiled by Statistics Netherlands

- Although mobile phone metadata are a rich source of information, it is clear that this kind of research is still in its infancy and requires a different approach than that of traditional data collection and analysis (e.g. surveys and administrative data sources). Many challenges still remain, not only in the field of tourism and mobility in particular, but also more generally in terms of processing and analysing big data sources. In the case of tourism, for example, one question is how many tourists does one mobile device represent? Foreign tourists may switch off their mobile phone, a family may carry only one mobile phone, while business travellers may have two mobile phones. Tourists from outside Europe may buy a SIM card in the Netherlands, which would result in underestimation of the number of foreign tourists for specific groups. And also the cost of calling abroad can play a different role for different groups of tourists. The representativeness of the data is therefore an issue. One obvious option would be to compare the number of devices at different events or areas with each other, to show the relative importance of these events or areas (benchmarking).

- Protecting privacy is extremely important when using this kind of data, and indeed their very use is sensitive to public opinion. Several measures were put in place in this pilot to guarantee confidentiality. First, all micro data remained with the telecom provider (Vodafone). Requests for data from SN where dealt with by an intermediary organization (Mezuro). Second, SN only received aggregated data and each area in the aggregated dataset contained at least 15 devices in a certain

timeframe. If there were fewer than 15 devices in an area, no data were supplied to SN. Third, mobility patterns of anonymised roaming visitors could only be analysed for one day. The following day the same phone would be assigned a new ID to prevent any risk of indirect identification if source data were to be disclosed unintentionally. And lastly, data were processed in a secure environment, only accessible for authorised researchers, and audited for this purpose by independent research organisation TNO.

- Results of measurements based on big data sources also require new methods of data visualisation, such as geo-location maps with time indicators. The challenge here is to produce visualisations that can be clearly understood and interpreted by users, such as policy-makers.

The overall conclusion of this pilot is, therefore, that data based on Call Detail Records and unique foreign mobile callers seem to have wide and feasible potential for tourism statistics, especially when they provide more details in terms of in place and time. It is possible to, for example, zoom in on small specific tourist areas and very short timeframes, such as certain events, coastal areas, or national holidays; to distinguish data per country of residence; or determine the number of transit passengers.

The potential for tourism statistics could be enlarged considerably if a mobile device could be followed for longer than one day. In a longer term development perspective, this would make it possible to distinguish between numbers, distances and frequency of daytrips, short and long holidays, and possibly derive modes of transport. Furthermore, it would also make it possible to distinguish flows of tourists, i.e. groups of tourists could be followed from the time they enter the Netherlands right up until they depart. We could then also determine which tourism locations are inter-related, even distinguished by the country of residence. This would generate new and important information for the tourism industry and policymakers that cannot be produced with traditional data collection and analysis. Similar methods could also be applied for outbound and domestic tourism, although for the latter case, the subjective definition of tourism – people spending time outside their usual environment – would be mean more of an obstacle; tourism would have to be objectified by variables like duration, distance and frequency. Future profiling of 'the usual environment' may be able to solve this issue: calls could be analysed to determine a person's usual environment by taking into account where they live, work, sport or go to school. Lastly, mobile phone metadata can also be combined with other data such as the effects of weather, the supply and quality of the accommodations, number of events, and the economic situation of a region or country of residence.

4. Final conclusions

The existence of new "big" data sources, like the internet and social media, mobile phone data, but also other sources like public transport data, data from electronic traffic loops, cameras and credit card data, may change the production landscape of tourism statistics significantly in the future. Not only will it affect the way existing tourism statistics are produced, it will also present tourism with new statistics which cannot be produced with traditional data collection and analysis. It won't happen tomorrow: there are still many challenges ahead, but the first results are promising. Data processing, methodology (e.g. representativeness) and, of course, privacy will remain important issues.

In essence big data sources have the potential to improve existing tourism statistics (better quality), to distinguish smaller areas and timeframes (more detail), to increase frequency and timeliness, to reduce the survey burden and produce new statistics and indicators (e.g. tourist flows) for policy-makers and the tourism industry. It is not technology and data availability that are putting the brakes on, but the development of methodology, the lack of big data knowledge and resources, privacy concerns, and the prevailing culture of statistics production.

The emergence of these new big data sources is comparable with the emergence of administrative data sources at the end of the last century. This also took a long time to sink in as an accepted and standard way of using data and making statistics. However, with big data there are differences. The internet is available for everybody, not only for statistical offices (democratisation of data). On the other side, companies who own big data sources are becoming more and more aware of the value of their data (commercialisation of data).

This affects the current role of national statistical institutes (NSIs). In this changing data landscape with new players and fields of force, they are slowly losing their relative monopoly on data and statistics. They will

find themselves in a more competitive environment, which is further enforced by the trend for open or public data. Therefore, NSIs will probably need to reposition themselves as the objective and unbiased party for high-quality reliable statistics for society as a whole.

5. References

1. Laney, D., 2001, 3D Management: controlling data volume, velocity and variety, Meta Group/Gartner;

2. Mayer-Schonberger, V. and K. Cukier, 2013, Big data: a revolution that will transform how we live, work and think, Eamon Dolan-Houghton Mifflin Harcourt;

3. Bouwman, H., N. Heerschap and M. de Reuver, 2013, Mobile handset study 2012: evaluation of mobile measurement software to monitor the use of smartphones and tablets in the Netherlands, TU Delft – CBS, Delft – The Hague, Netherlands;

4. Arends, J., M. Morren, F.Y. Wong and M. Roos, 2013, Smarter 2012, CBS, Statistics Netherlands, The Hague, Netherlands;

5. Tiru, M. and R. Ahas, 2013, Passive anonymous positioning data for tourism statistics, University of Tartu, Estonia;

6. Offermans, M., A. Priem and M. Tennekes, 2013, Rapportage project Impact ICT; mobiele telefonie, 16 April 2013, Statistics Netherlands, The Hague, Netherlands.