

## 14<sup>th</sup> Global Forum on Tourism Statistics

23-25 November 2016, Venice, Italy

### How many SIM in your luggage? A strategy to make mobile phone data usable in tourism statistics

Authors:

B. Dattilo

[dattilo@istat.it](mailto:dattilo@istat.it)

R. Radini

[radini@istat.it](mailto:radini@istat.it)

M. Sabato

[sabato@istat.it](mailto:sabato@istat.it)

ISTAT, Italian National Institute of Statistics

## **SESSION 1: MOBILE PHONE DATA FOR TOURISM STATISTICS**

### **ABSTRACT<sup>1</sup>**

*Big data* and, in particular, *mobile positioning data* represents an original source of information about the territorial movements of individuals and tourism flows. In the near future, the systematic use of this data for statistical purposes will provide various advantages in terms of cost-efficiency, timeliness, completeness and reduction of the “statistical burden”.

At present, the data obtained from mobile phone operators cannot replace, in full, information derived by household surveys currently carried out to measure tourism demand. In addition to the well-known legal, technological and financial problems, there are conceptual issues related both to the international definitions adopted in the field of tourism statistics and to methodological limitations which could compromise the access and use of mobile phone sources.

For example, data recorded by mobile phone providers, so-called CALL DETAIL RECORDS (CDRs) and DATA DETAIL RECORDS (DDRs), have problems concerning access and representativeness. The first one regards agreements with providers and the process of protecting anonymity of data adopted by them; the second one concerns undercoverage or overcoverage of the CDRs frame as compared to the target population.

However, a lot of experimentation carried out by some European countries, together with the promotion of the knowledge on this topic by Eurostat, have made evident, also in Italy, the need to explore the potential of mobile phone data. For this purpose, within the Italian National Statistical Program, a project about the use of mobile positioning data, has been launched. The aim is the production of indicators on domestic and inbound tourism at territorial and temporal level more detailed than those that can be estimated by a sample survey. The objective is to verify the feasibility of including these indicators in the official system of tourism statistics, in order to improve and complement traditional sources.

This paper presents the first results achieved by “Trips and Holidays” Survey, in which information about the relation between trips and the use of SIM during the trip has been collected since 2015. In particular, information concerning the number of “travelling” SIM, how SIM cards have been used (calls, SMS/MMS, internet connection) and the type of SIM owners. Data collected in this way will allow calculating calibrators to be applied to mobile phone data in order to estimate the tourism flows correctly.

Therefore, this study is an important first step towards the development of a methodology on big data use for statistical purposes, which will also have positive implications in other analysis and research contexts such as migration and demographic statistics.

**Keywords:** *big data, mobile positioning data, new sources*

---

<sup>1</sup> The paper reflects the joint efforts of the authors and their own opinions.

## **INDEX**

Introduction .....	4
1. The methodological framework .....	5
2. Using mobile positioning data in tourism statistics: the <i>art-of-the-state</i> in ISTAT .....	7
2.1. Mobile phone data: constraints and opportunities .....	7
2.2. The project ISTAT .....	8
3. SIM and trips: first results.....	9
Conclusions .....	14
References .....	15

## Introduction

Tourism represents one of the best domains of application to use *mobile positioning data* for statistical purposes. The data collection from Mobile Network Operators (MNOs) would enable to have a huge potential of information about the territorial movements of individuals, in general, and tourism flows, in particular.

At present, data obtained from this source cannot meet the information needs expressed by the legal framework for European statistics on tourism (Regulation 692/2011). However, the advantages that a systematic use of these data could ensure in the future (in terms of cost-efficiency, timeliness, accuracy, reduction of “response burden”) push to explore the potential of mobile phone data through ad hoc experimentations.

The presentation of a methodology to use anonymized mobile phone data for tourism statistics is the subject of this work. This methodology has been developed within the project “Experimentation on the use of mobile phone data for statistics on tourism demand”, launched by ISTAT and included within the Italian National Statistical Program.

The aim of this project is the calculation of calibrators to be applied to the so-called Call Detail Records (CDRs) or Data Detail Records (DDRs) provided by MNOs, in order to “boil down” the call/connection events by mobile phone to tourism events. The construction of these calibrators is based on the information about the relation between trips and “travelling” SIM, collected by “Trips and Holidays Survey” (THS) since 2015.

The assumption is that the information integration achieved by two sources - data from the official statistical system (THS) and data from mobile phone traffic (CDRs/DDRs) - could, in general, improve the quality of the tourism flows estimates and, in particular, allow the production of indicators on tourism at territorial and temporal level more detailed than those that can be estimated by a sample survey.

The first part of the paper describes the Istat’s *roadmap* for the adoption of big data, the actions taken by the Institute, the projects already completed and the work in progress. Right after, the focus is on the mobile positioning data, their features and on the technological, legal and financial challenges to use them for statistical purposes.

The second part of the paper analyses the conceptual and methodological issues related to the use of mobile positioning data in the specific field of tourism statistics; at the same time, the need to explore the potential of this source is emphasized, also taking into consideration of several experimentations already carried out by some European countries, together with the promotion of the knowledge on this topic by Eurostat.

The work continues describing the above-mentioned project conducted in the THS and included in the National Statistical Program, its short, medium and long term objectives and the methodology adopted.

The last part of the paper is dedicated to the illustration of the preliminary results achieved by the THS about the relation between trips and “travelling” SIM in 2015, focusing on the potential use of these information assets, in order to test in the near future the CDRs/DDRs to produce estimates on tourism.

## 1. The methodological framework

Daily use of information and communication technologies, such as mobile phones, creates a trail of digital breadcrumbs that people spread around during their movements all the time. In particular, such digital breadcrumbs include mobile positioning data, which represent a relevant source of information about the territorial movements of individuals and tourism flows.

These data are recorded by mobile phone providers as Call Detail Records (CDR); CDRs include incoming and outgoing call activities, sending and receiving SMS and MMS messages. CDRs also include internet traffic between the mobile devices and the network; such traffic is often referred as DDR (Data Detail Record: record of internet data traffic). A DDR is a part of a CDR. Mobile devices with internet connection turned off or without the internet capability do not produce any DDR. Conversely, when internet capabilities are enabled, a huge number of DDRs is produced, because for instance many apps installed on smartphones actively use internet connection to exchange information even when the phone is in the pocket.<sup>2</sup>

CDRs are described by: User identifier, Time of activity (e.g. when the call was made) and Location of activity as Cell (e.g. antenna ID, where the call was made).

There are some main obstacles for accessing and using these data, namely: 1) privacy and confidentiality of data (legal requirements); 2) business models, with potential market interest collision between MNOs and different business perspectives between MNOs and National Statistical Institutes (NSIs).

Regarding the first problem, many EU member states have a legal vacuum around data access legality. In Italy, there isn't any specific legislation protecting Mobile Phone data access. For their particular nature, the Mobile Phone Location Data can be referred to individuals, and are considered personal (as potentially sensible)<sup>3</sup> and must be treated in such a way to avoid disclosure.

In 2014, ISTAT submitted a request to the Italian Data Protection Authority (Privacy Guarantor) in order to obtain permission to receive mobile phone data from a National MNO (WIND) and the Privacy Guarantor recommended ISTAT to provide appropriate guarantees, i.e. maintaining collected data in an anonymous format and with no possibility of linkage between an anonymous code and the original identification<sup>4</sup>. This problem can be overcome by adopting anonymization procedures (already adopted by Wind) to prevent re-identification or indirect identification. The indirect identification of personal data may take place whenever it is possible to isolate some or all records which identify an individual in the data set (singling out of a record), or to link, at least, two records concerning the same individual in the same database or in two different databases ("linkability" of two records). The indirect identification of personal data is also possible when it can deduce, with significant probability, the value of an attribute from the values of a set of other attributes (inference of attributes).

Regarding the second issue – business models - the location where the data processing is taking place is crucial, because the MNOs have business secrets within the data and they want to make sure that only statistical office could access the data and nobody else (especially not other MNOs). There is the need to build collaborative partnerships that could be a "win-win" opportunity for both public and private interests.

ISTAT set up a technical Committee in February 2013 with the purpose of designing a strategy for Big Data adoption in Official Statistics. The workplan of the Committee identified three experimental projects in order to concretely test the possible adoption of Big Data in Official Statistics. Mobile phone data were used in the pilot projects "Estimating inter-city mobility using CDRs, in an integrated approach with administrative data" launched by the Committee. This experiment had the purpose to investigate to what extent mobile calls data, integrated with administrative data, could be used to produce reliable and timely estimates of inter-city mobility. This study was jointly developed by ISTAT, CNR, University of Pisa. In an ongoing project at ISTAT, called "**Persons and Places**", it was produced a first release of the Origin-Destination matrix related to movements across municipalities for the purpose of work and study; this matrix was produced starting mainly from administrative data. Within the joint experimental project, the purpose was to produce the same matrix starting from CDRs. The dataset used in this work consisted of CDRs collected in 5 weeks in the province of Pisa, Tuscany. As already mentioned the privacy and confidentiality of data were guaranteed by:

---

<sup>2</sup> Tiru M. (2014), *Overview of the sources and challenges of mobile positioning data for statistics*, Beijing – China: International Conference on Big Data for Official Statistics, 28-30 October 2014.

<http://unstats.un.org/unsd/trade/events/2014/beijing/Margus%20Tiru%20-%20Mobile%20Positioning%20Data%20Paper.pdf>

<sup>3</sup> Personal data are defined by Art. 4, paragraph 1, lett. B of Legislative Decree. N. 196/2003 and data anonymized and not indirect identification are defined by Art. 4, paragraph 1, letter. N of Legislative Decree N. 196/2003.

<sup>4</sup> Parere sul PSN 2014-2016 aggiornamento 2015-2016 - 18 settembre 2014.

1) anonymized CDRs; 2) restricting local information to municipality level (no cells); 3) time limitation to 5 weeks. The positive comparison of the Origin Destination matrix generated from administrative data with respect to the same matrix created from CDRs was an initial relevant confirmation of the potential usage of these kind of data for Official Statistics purposes<sup>5</sup>.

The further two experimental projects supervised by ISTAT Big Data Committee were: 1) **Google Trends**<sup>6</sup>, aimed at exploiting the time series of query shares extracted from Google Trends as auxiliary variables to improve the quality of estimates of short-term (monthly or quarterly) socio-economic indicators; 2) **Use of web scraping and text mining techniques in the ISTAT survey on “Information and Communication Technology in enterprises”**<sup>7</sup>, aimed at producing information on the use of ICT and in particular on the use of Internet by Italian enterprises for various purposes.

ISTAT Technical Committee on Big Data also examined the technological setting for collecting and processing large amounts of data for statistical purposes, coming up with a twofold strategy: 1) setting up an internal laboratory for experimenting and tuning new technologies and systems; 2) using an external laboratory for production-ready, resource-intensive applications (e.g. Sandbox of Big Data Project UNECE 2014<sup>8</sup>). The technological issues are not as relevant as the statistical ones. However, specific investment should be done for buying hardware and software platforms, and skilling IT people for their usage and maintenance. In addition, for mobile phone data, an additional cost depends on whether the MNO has to process data into aggregated statistical indicators before transmitting to NSIs (this elaboration could potentially resolve privacy problem) or they could just extract and transmit raw data to NSIs for processing. The overall cost for an NSI is obviously higher in the former case than compared to the latter<sup>9</sup>.

ISTAT participates to the project ESSnet Big Data since February 2016. This is a project within the European statistical system (ESS) jointly undertaken by 22 partners. Its objective is the introduction of Big Data within the regular production of Official Statistics, through pilots studies exploring the potential of selected Big Data sources and building concrete applications. The project is composed by 10 workpackages: eight of these (WP1 to WP8) are content-oriented, while the other two support the overall project. The workpackages focus on webscraping of job vacancies and enterprise characteristics, mobile phone data, smart meters, and more particularly on electricity meters, AIS data, and methodology for processing data resulting from a combination of big data sources, administrative data and statistical survey data. So far, the workpackage “mobile phone data” is investigating the possibility to access mobile phone data from institutional and legal (privacy) point of view for several of the 22 partners. In particular, a survey was launched to study the legal constraints and the level of individual negotiation between partners and national MNOs. In a second phase the workpackage will conduct pilot studies demonstrating use cases (also for the statistics on tourism) for producing or enriching official statistics.

A further problem related to the use of mobile phone data for Official Statistics concerns the communication to the general public of the use of the telephone data as an additional source. The key players (ISTAT, Privacy Guarantor, and MNOs) have to be very clear in the explanations to the general public, clarify that this use shouldn't be perceived by the public as a means for having a control or tracking or spying on the personal habits of individuals. The purpose is instead to compile collective, anonymous statistics concerning the phenomena of tourism activities (or another relevant domain)<sup>10</sup>.

---

<sup>5</sup> Furletti B., Gabrielli L., Garofalo G., Giannotti F., Milli, L., Nanni M., Pedreschi D., Vivi, R. (2014), *Use of mobile phone data to estimate mobility flows. Measuring urban population and inter-city mobility using big data in an integrated approach.*, Cagliari – Italy: 47th Scientific Meeting of the Italian Statistical Society 2014, 11-13 June 2014.

<sup>6</sup> Bacchini, F. , D'Alò M., Falorsi S., Fasulo A., Pappalardo A. (2014), *Does Google index improve the forecast of Italian labour market?*, Cagliari – Italy: 47th Scientific Meeting of the Italian Statistical Society 2014, 11-13 June 2014.

<sup>7</sup> Barcaroli G., Nurra, A., Scarnò M., Summa D. (2014), *Use of web scraping and text mining techniques in the ISTAT survey on “Information and Communication Technology in enterprises*, Wien –Austria: European Conference on Quality in Official Statistics (Q2014) , 3-5 June 2014.

<sup>8</sup> Link to public-access wiki space for UNECE work relating to Big Data <http://www1.unece.org/stat/platform/display/bigdata/Big+Data+in+Official+Statistics>

<sup>9</sup> Baldacci E., Scannapieco M., edit by, (2015), *ISTAT's Roadmap for the Adoption of Big Data for Official Statistics*, 20 April 2015.

<sup>10</sup> IFSTTAR, NIT, Positium LBS, Statistics Estonia, Statistics Finland, University of Tartu, (2014), *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*, Eurostat Contract No. 30501.2012.001-2012.452, voll 1-4, <http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies>.

## 2. Using mobile positioning data in tourism statistics: the *art-of-the-state* in ISTAT

### 2.1. Mobile phone data: constraints and opportunities

The previous paragraph has pointed out the main barriers (privacy and regulatory issues, financial and business-related problems, technological issues) that must be overcome to access and use mobile positioning data for statistical purposes.

The mobile phone data processing, starting from CDRs/DDRs to produce tourism statistics, requires not only the removal of these obstacles, but also the resolution of both conceptual issues related to international definitions adopted in the field of tourism statistics, both methodological problems of representativeness.

CDRs/DDRs include data on call/internet connection via mobile events, that allow to identify the *real time location* of an individual; this would enable to track the movements of individuals in time and space, including tourism flows.

Therefore, a crucial aspect is the need to correctly define the target population, distinguishing tourism and non-tourism activities. This mainly involves the accurate identification of the place of residence and, accordingly, of the usual environment of subscribers, in order to differentiate between regular trips and tourism trips, between commuters and tourists, etc.

The coverage is another problematic aspect of representativeness of mobile positioning data: the CDRs frame suffers from problems both of undercoverage due, for example, to the fact that not all tourists use mobile phones *tout court* or during a trip, both of overcoverage caused by the fact that some tourists use more than one device and/or SIM cards of which they are not direct holders. Hence there is the need to carefully evaluate and calibrate the information of the CDRs/DDRs, also using other data sources.

The validity and accuracy of mobile positioning data have been widely discussed in the feasibility study<sup>11</sup> that Eurostat conducted between 2012 and 2014 to promote the knowledge and the use of mobile phone data within the framework of tourism statistics. On that occasion, several issues related to the use of the mobile phone source were discussed: the data accessibility from a legal, technological, financial and business point of view, the methodological and conceptual aspects, the comparison with existing sources, the limits and the opportunities.

Furthermore, the study indicated the availability of the longitudinal mobile phone data as the best methodological option for compiling tourism statistics: *“Longitudinal data is a must for reliable tourism statistics in order to assess the whereabouts of the subscribers over a longer period of time [...]”*

*The use of non-longitudinal initial data [...] limits the possibilities of generating reliable statistical indicators drastically, mainly because it is impossible to measure the duration (and frequency) of the stays of subscribers in specific places, which is crucial for identification of a subscriber’s country of residence and usual environment”<sup>12</sup>.*

One of the conclusions of the study is that, at present, data by MNOs are not able to provide some information required by the current Regulation 692/2011 concerning European statistics on tourism (purpose of the trip, expenditure, type of accommodation and means of transport used, etc.). However, mobile positioning data can be rather a supplementary source for the traditional ones, because they introduce important elements of improvement in terms of cost-efficiency, timeliness (in some cases up to real time), potential access to new information, completeness, reduction of the “statistical burden”, better accuracy in time and space.

Moreover, the technological evolution is very fast and from the increasing availability of information, it could allow to suppose a scenario in which the traditional data sources on tourism will be completely replaced by mobile phone data. In this sense, Eurostat encourages NSIs to quickly acquire technological and methodological know-how to face the challenges and catch the new opportunities offered by big data.

After all, in recent years, at European level, several experimental initiatives on the use of mobile positioning data for tourism statistics have been carried out; some European countries have used them for the production of official statistics as Estonia<sup>13</sup>, Czech Republic<sup>14</sup> and Netherlands<sup>15</sup>. Other countries, as

<sup>11</sup> IFSTTAR, NIT, Positium LBS, Statistics Estonia, Statistics Finland, University of Tartu, (2014), *op. cit.*

<sup>12</sup> IFSTTAR, NIT, Positium LBS, Statistics Estonia, Statistics Finland, University of Tartu, (2014), *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics. Consolidated Report*, Eurostat Contract No 30501.2012.001-2012.452, pp 5-7.

<sup>13</sup> Ahas R., Aasa A., Silm S., Tiru M. (2007), *Mobile positioning data in tourism studies and monitoring: case study in Tartu, Estonia*, in Sigala, M., Mich, L., Murphy, J. (Eds.), *Springer Computer Science: Information and Communication Technologies in Tourism*, pp. 119-

Ireland, Slovenia, Montenegro and Finland, have embarked to a path to obtain data from MNOs; others (Portugal, Austria, Switzerland, the United Kingdom) have used them, to perform special analysis on the tourism field<sup>16</sup>.

## 2.2. The project ISTAT

Recently also in Italy, the need to explore the potential of mobile phone data has emerged. For this reason, together with the solution of the knotty problems of accessibility/use of mobile phone data from a legal, financial and technological perspective, and with the establishment of agreements with MNOs for CDRs/DDR provision, the project work "Experimentation on the use of mobile phone data for statistics on tourism demand" has been launched.

This initiative promoted by ISTAT and included within the Italian National Statistical Program, is part of the above-mentioned roadmap for the Adoption of Big Data for Official Statistics<sup>17</sup>, which is a product of the work of the technical Committee set up by the Institute in February 2013.

The ultimate aim of this study is the access and the use of mobile positioning data in order to obtain information about tourism flows by improving the timeliness, the completeness and, in general, the quality of data attained with the traditional statistical sources on tourism.

In particular, the **short and medium term objectives** of this work are: 1) definition and experimentation of a methodology for the use of mobile phone data aimed at the construction of *calibrators* to be applied to the CDRs/DDRs, in order to overcome the methodological limitations, especially coverage; 2) testing and assessment of the results obtained in terms of validity and accuracy, even by means of a benchmark with traditional sources.

The **long-term objectives** are: 1) the production of indicators on domestic tourism more timely and disaggregated than those that can be estimated by the current survey on tourism demand. These indicators would be helpful, as a benchmark, also for ISTAT supply side statistics; 2) the spread of the methodology to product indicators on inbound tourism, actually provided by ISTAT supply side statistics (for the part of tourist accommodation establishments) and by Bank of Italy leading the border survey, at more detailed territorial or temporal level.

The first phase of the study, the results of which will be discussed in the next paragraph, has concerned the setup of an original list of questions about the relation between trips and the use of SIM during the trip. New questions have been included within the questionnaire of THS since 2015.

In particular, the information collected concerns:

- the use of the mobile phone during the trip;
- how to use the mobile phone during the trip (calls, SMS / MMS, internet connection);
- the number of "travelling" SIM;
- the type of SIM owners (participants/non participant in the trip, companies, etc.).

The information as collected allows, firstly, to assess the reliability and the level of CDRs/DDRs coverage regards to the domestic tourism phenomenon and, secondly, to calculate calibrators to be applied to the mobile phone data in order to estimate tourism flows.

Of course, this method can only be applied if some preliminary operations of "normalization", regarding the usual environment in temporal and spatial terms, have been already carried out over the CDRs/DDRs dataset.

---

128. Ahas R., Silm S., Järv, O., Saluveer E., Tiru M. (2010), *Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones*, Journal of Urban Technology, 17(1): 3-27. [http://statistika.eestipank.ee/failid/mbo/valisreisid\\_eng.html](http://statistika.eestipank.ee/failid/mbo/valisreisid_eng.html).

<sup>14</sup> Jonge, E., Pelt, M., Roos, M. (2012), *Time patterns, geospatial clustering and mobility statistics based on mobile phone network data* Statistics Netherlands, Division Methodology and Quality. The Hague/Heerlen <http://www.cbs.nl/NR/rdonlyres/010F11EC-AF2F-4138-8201-2583D461D2B6/0/201214x10pub.pdf>.

<sup>15</sup> Vogelová, M. (2012), *Using residual positioning data from mobile networks for tourism monitoring*, Reykjavik – Iceland: 11th Global Forum on Tourism Statistics, 14-16 November 2012. [http://www.congress.is/11thtourismstatisticsforum/presentations/Marketa\\_Vogelova.pdf](http://www.congress.is/11thtourismstatisticsforum/presentations/Marketa_Vogelova.pdf).

<sup>16</sup> Positium LBS. (2014), *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*, Eurostat contract no. 30501.2012.001-2012.452 Task 1. Stock-taking.

<sup>17</sup> Baldacci E., Scannapieco M., edit by, (2015), *op.cit.*

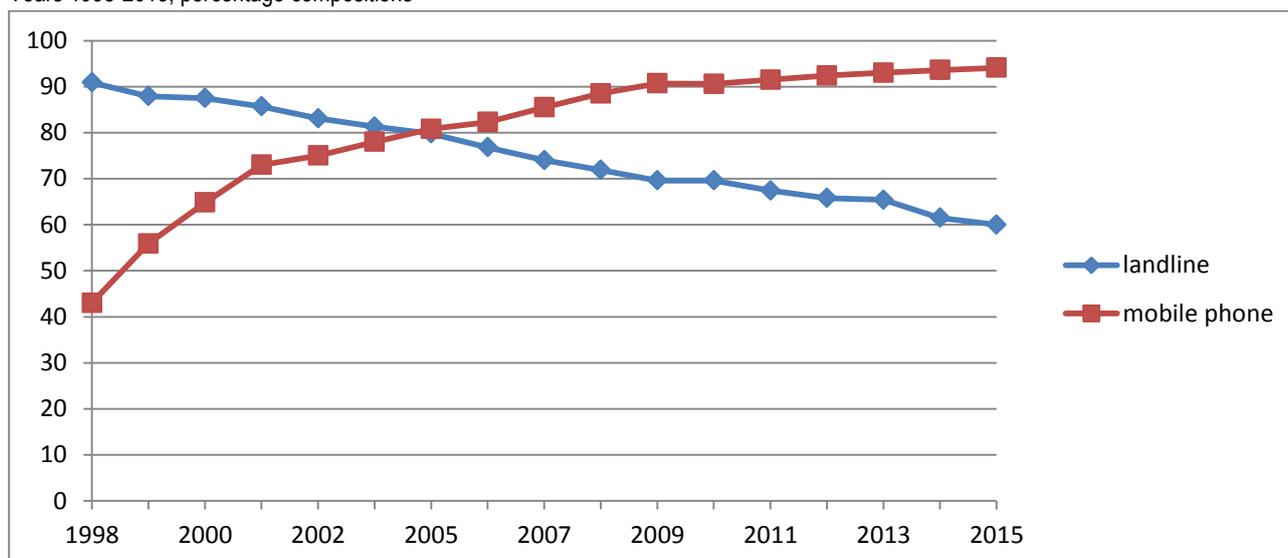
### 3. SIM and trips: first results

Since 2014, THS has abandoned the CATI technique opting for the CAPI technique in order to overcome the undercoverage problem, which stressed the sampling frame (the national list of fixed line telephone numbers). When THS was built (in 1997), the factor which made the use of CATI possible was the widespread availability of landline phone in houses: in fact, households with listed fixed phone number were more than 85%, the only possession of the mobile phone was still marginal (just over 3%), and the number of households without any phone (about 6%) was gradually decreasing. This situation had gradually become less favorable, mainly due to the sudden spread of mobile phones, which had frequently replaced the fixed ones, and to the growth of subscribers who had chosen to have an unlisted phone number. In 2012, the percentage of households out of reach by means of the list amounted to 51.7%, representing for the first time the majority of households<sup>18</sup>.

On the one hand, the significant spread of mobile phone was the constraint that had influenced the decision to change survey technique, on the other hand it could be an opportunity for the development of future analyses, especially in light of the fact that mobile positioning data currently have an almost total coverage. In 1998, 90.9% of households had a landline (reserved or not), only 43% had at least one mobile phone. Since then, this percentage had been steadily increasing, meanwhile the amount of households with fixed line phones had been rapidly descending. In 2005, there was the overtaking: the percentage of households with a mobile phone (80.8%) exceeded that of households with fixed phones (79.8%). In 2015, almost 24 million households, amounting to 94.1% of households resident in Italy, owned a mobile phone and only 60% a fixed one (Figure 1).

FIGURE 1 – HOUSEHOLDS PER POSSESSION OF THE PHONE.

Years 1998-2015, percentage compositions



Source: ISTAT- Multipurpose Survey on households: aspects of daily life

In order to assess the reliability of the information included in the CDRs/DDR and, most importantly, their level of coverage, it is necessary to have available data on *how much* and *how to use* the mobile phone during tourism movements.

In 2015, the mobile phone was used in 86,1% of trips to call and in 71% of trips to send SMS or MMS; these percentages were similar both during holidays both during business trips. The use of the mobile phone for internet connection was higher in business trips than in holidays (68.7% of business trips against 57.4% of holidays for SIM connections, 64.7% against 43.6 % for Wi-Fi connections). Taking into consideration the destination, the percentage of trips where the mobile phone was used is greater in domestic trips, regardless

<sup>18</sup> Di Torrice M., Morricono M., Sabato M. (2012), *Beyond the CATI technique: a benchmarking analysis on demand-side surveys in EU and the Italian choice*, Reykjavik – Iceland: 11th Global Forum on Tourism Statistics, 14-16 November 2012.

of the type of use (understandable since there is the international roaming abroad), except for the internet connection via Wi-Fi (free and more widespread abroad) (Table 1).

**TABLE 1. TRIPS USING MOBILE PHONE PER TYPE OF USE, TYPE OF TRIP AND MAIN DESTINATION.** Year 2015, percentage compositions

MOBILE PHONE USE	TYPE OF TRIP			DESTINATION		
	HOLIDAY TRIPS	BUSINESS TRIPS	ALL TRIPS	DOMESTIC	OUTBOUND	ALL TRIPS
call	85,9	87,6	86,1	89,6	71,3	86,1
send sms or mms	71,1	70,3	71,0	73,0	62,6	71,0
internet with SIM	57,4	68,7	58,8	61,6	46,9	58,8
internet with Wi-Fi	43,6	64,7	46,1	43,4	57,7	46,1

Source: ISTAT – Trips and holidays Survey

Only the first three types of mobile phone use (calls, send SMS or MMS, internet connect by SIM) produce a voice/data traffic that is traceable by CDRs/DDRs and then available to estimate tourism flows. In fact, in this case, it is possible to talk about events related to the use of a SIM card. This means that, in order to assess the level of coverage of the CDRs/DDRs, it is necessary to consider the distribution of trips by the use of at least one SIM card during the trip.

In 2015, the percentage of trips where at least one SIM card - to call, send SMS / MMS, connect to the Internet - was used, was 88.2% of trips (88.1% of holidays, 88.5% of business trips).

It is assumed that the CDRs/DDRs, which MNOs will provide, only include data that track the "telephone events" of SIM registered in Italy and assigned to Italian individuals/companies; so, the following analysis of the data collected by THS only concerns the domestic tourism. Taking into account only this part of the tourism of residents, in 2015, the percentage of trips, where at least one SIM was used, rose to 90%. Out of the observation field of the CDRs/DDRs, therefore, there was about 12% of trips in general, and 10% of domestic trips, in particular (Table 2). The undercoverage problem of the SIM could be overcome by designing a calibration system based exactly on the collection of information by THS.

**TABLE 2. TRIPS PER SIM USE, TYPE OF TRIP AND MAIN DESTINATION.** Year 2015, absolute values in thousands and percentage compositions.

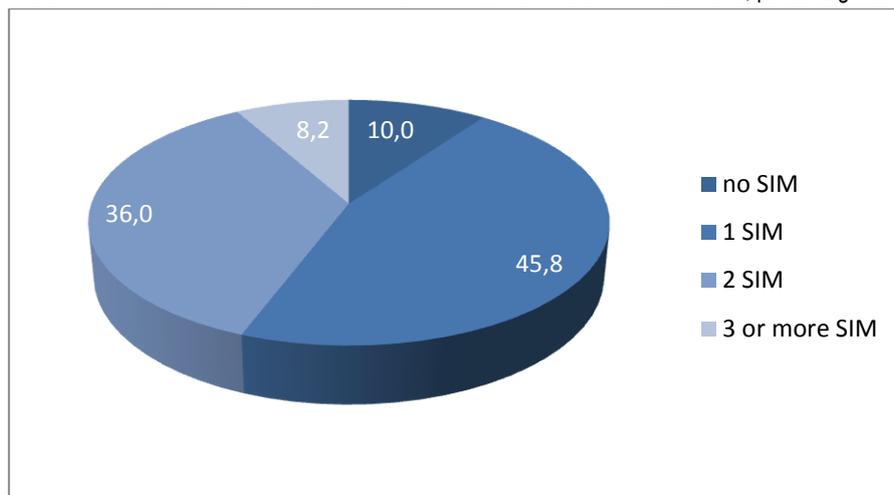
SIM USE DURING THE TRIP	TYPE OF TRIP						DESTINATION					
	HOLIDAY TRIPS		BUSINESS TRIPS		ALL TRIPS		DOMESTIC		OUTBOUND		ALL TRIPS	
	ths	%	ths	%	ths	%	ths	%	ths	%	ths	%
YES	45.129	88,1	6.103	88,5	51.232	88,2	42.403	90,0	8.829	80,1	51.232	88,2
NO	6.092	11,9	791	11,5	6.884	11,8	4.690	10,0	2.194	19,9	6.884	11,8
<b>TOTAL</b>	<b>51.222</b>	<b>100,0</b>	<b>6.894</b>	<b>100,0</b>	<b>58.115</b>	<b>100,0</b>	<b>47.093</b>	<b>100,0</b>	<b>11.023</b>	<b>100,0</b>	<b>58.115</b>	<b>100,0</b>

Source: ISTAT – Trips and holidays Survey

Among domestic trips where at least one SIM was used, 45.8% was characterized by the use of only one SIM, 36% by two SIM, 8.2% by three or more SIM (Figures 2 ). In 2015, the average number of participants in the same domestic trip was 1.6, while the average number of SIM per capita during a trip amounted to 0.8<sup>19</sup>.

<sup>19</sup> The average number per capita of travelling SIM was calculated by dividing the total number of travelling SIM in Italy during the reference year to the total number of participants in domestic trips where at least one SIM was used.

FIGURE 2 – DOMESTIC TRIPS PER NUMBER OF SIM USED . Year 2015, percentage compositions.



Source: ISTAT – Trips and holidays Survey

It should be highlighted that behind each trip may be a “bunch” of people travelling together; therefore, the number of travelling SIM is closely related to the number of participants in the same trip. In 2015, when one individual travelled in Italy, in 93.9% of trips, just one SIM travelled with him; the percentages of trips where only one tourist brought two SIM (5.9%) or more (0.2%) were residual. When two members of the same household travelled together, in the majority of cases, there was a match between the number of participants and the number of travelling SIM (in 61.9% of trips with two participants, two SIM cards travelled, in 34.9%, one single SIM and in 3.2% three or more SIM). When three or more components of the same household had a trip, the relation between number of travelling SIM and number of participants was reversed: in only 23.6% of trips there was a correspondence between SIM number and number of participants while, in the majority of trips, the number of travelling SIM was lower than the number of the participants (56.6%) and in the remaining 19.9% was higher (Table 3). This could be explained by the fact that with the increase of the number of participants in the same trip, the probability that other members, who travelled together, are also children (who normally do not have mobile phone or are not holders of a SIM) grows.

TABLE 3. DOMESTIC TRIPS PER NUMBER OF SIM USED AND NUMBER OF PARTICIPANTS IN THE SAME TRIP. Year 2015, absolute values in thousands and percentage compositions (per 100 trips with the same characteristics)

NUMBER OF SIM	NUMBER OF PARTICIPANTS							
	1		2		3 or more		TOTAL	
	ths	%	ths	%	ths	%	ths	%
1	14.282	93,9	4.392	34,9	2.902	19,9	21.576	50,9
2	900	5,9	7.791	61,9	8.265	56,6	16.956	40,0
3 or more	25	0,2	401	3,2	3.445	23,6	3.871	9,1
<b>TOTAL</b>	<b>15.207</b>	<b>100,0</b>	<b>12.584</b>	<b>100,0</b>	<b>14.612</b>	<b>100,0</b>	<b>42.403</b>	<b>100,0</b>

Source: ISTAT – Trips and holidays Survey

If the perspective shifts from trips to travelling SIM, simulating the processing of the CDRs/DDR information asset which it should tend towards, it can observe the distribution of the number of participants in the same trip by the number of SIM card used during the trip. In 2015, for 70.8% of travelling SIM in Italy, there was a perfect correspondence between the SIM number and the number of participants in the trip, while, in the remaining 29.2% of cases there was a mismatch. In particular, when only one SIM travelled, in 12.6% of cases there were two participants and only in 4.9% three or more. When two SIM travelled, in a significant part (32.2%), there were three or more members of the same household travelling together. Finally, when the tourist brought three or more SIM cards, in 24.5% of cases there was a lower number of participants in the same trip (Table 4).

**TABLE 4. TRAVELLING SIM IN ITALY PER NUMBER OF PARTICIPANTS IN THE SAME TRIP.** Year 2015, absolute values in thousands and percentage compositions on travelling SIM in Italy

NUMBER OF PARTICIPANTS	NUMBER OF SIM							
	1		2		3 or more		TOTAL	
	ths	%	ths	%	ths	%	ths	%
1	14.282	82,5	1.801	12,6	76	1,8	16.158	45,1
2	2.172	12,6	7.866	55,2	961	22,7	10.998	30,7
3 or more	851	4,9	4.584	32,2	3.200	75,5	8.634	24,1
<b>TOTAL</b>	<b>17.304</b>	<b>100,0</b>	<b>14.250</b>	<b>100,0</b>	<b>4.236</b>	<b>100,0</b>	<b>35.790</b>	<b>100,0</b>

Source: ISTAT – Trips and holidays Survey

As is well known, there is not always a match between the SIM holder and the individual who used it during the trip. Data reported in the following table may be useful to typify the information in the CDRs/DDR. In almost all cases (96.4%), travelling SIM were assigned to members of the household participating in the same trip; 2.6% of travelling SIM were owned by persons not belonging to the household or assigned to companies (Table 5).

**TABLE 5. TRAVELLING SIM IN ITALY PER TYPE OF SIM OWNER.** Year 2015, absolute values in thousands and percentage compositions on travelling SIM in Italy

TYPE OF SIM OWNER	SIM	
	ths	%
member of the household participating in the same trip	34.485	96,4
member of the household NOT participating in the same trip	232	0,6
person not belonging to the household or company	948	2,6
does not know/does not remember	125	0,3
<b>TOTALE</b>	<b>35.790</b>	<b>100,0</b>

Source: ISTAT – Trips and holidays Survey

Data collected by THS and described in this paper give information of reference scenario, supporting those that may be arisen from the CDRs/DDR. This information asset allows to draw the lines of the context in which we are moving.

First of all, these data enable to assess the level of coverage and the reliability of the CDRs/DDR. Wanting to go further, the same information could be used to calculate calibrators to be applied to the CDRs/DDR, in order to estimate the amount of domestic tourism.

Therefore, in 2015, 35 million and 790 thousand travelling SIM in Italy were used in 47 million and 93 thousand domestic trips; their ratio, amounting to 1.3, is the calibrator to be applied to the CDRs/DDR. In other words, using the information from THS, which measures both domestic trips both travelling SIM in Italy, it could compare the two amounts to calibrate the information asset of the CDRs/DDR for the estimates of tourism flows, overcoming, in fact, the undercoverage problem of mobile positioning data (which amounted, in 2015, around 10% of domestic trips).

Fine-tuning the calibrator and taking into account the tourist geographical area of residence, it gets the results shown in the table below (Table 6).

In 2015, in the Northwest and Insular Italy, for example, the calibrator to be applied to the CDRs/DDR was higher (1.6 and 1.5 respectively) than the national average because the number of trips where any SIM card was used was greater than in other areas of the country. In the Northeast and in the Centre Italy, the calibrator was lower (1.2 in both areas), because the relation between travelling SIM and trips was more balanced.

**TABLE 6. TRAVELLING SIM AND DOMESTIC TRIPS PER GEOGRAPHICAL AREA OF RESIDENCE.** Year 2015, absolute values in thousands and percentage compositions

GEOGRAPHICAL AREA OF RESIDENCE	TRAVELLING SIM IN ITALY		DOMESTIC TRIPS		Calibrators to be applied to CdRs/DDRs
	ths	%	ths	%	
North-West	7.138	19,9	11.365	24,1	1,6
North-East	12.757	35,6	15.633	33,2	1,2
Centre	8.755	24,5	10.518	22,3	1,2
South	5.770	16,1	7.587	16,1	1,3
Islands	1.371	3,8	1.989	4,2	1,5
<b>ITALY</b>	<b>35.790</b>	<b>100,0</b>	<b>47.093</b>	<b>100,0</b>	<b>1,3</b>

Source: ISTAT – Trips and holidays Survey

## Conclusions

In the previous pages, preliminary results of the ISTAT project on mobile positioning data for tourism statistics has been presented; next steps will include the improvement and the application of the methodology implemented on a test dataset of the CDRs/DDR, provided by one or more MNOs.

This is about assessing the potential and the reliability of calibrators and, therefore, the goodness of the methodological approach to mobile positioning data proposed here. The way indicated in this work is only one of the possible and different means to verify the feasibility of including mobile data in the official system of tourism statistics, with the aim to complement and support traditional sources.

The proceeded approach involves the use, as a benchmark for the estimates obtained with the mobile phone data, of the results achieved by THS. However, in order to ensure the activities of the experimentation will go ahead, it is necessary, firstly, to overcome some methodological obstacles related to the definition of the observation period of the "phone events", also functional to the usual environment delimitation of the mobile phone subscriber.

Secondly, a fundamental step is the conclusion of agreements with MNOs about data provision.

In fact, this is about:

- defining which types of data MNOs will provide, only CDRs in the strict sense or even DDRs. The decision will influence the calculation of calibrators that will be eventually adjusted, in order to take into account or not the use of travelling SIM for the Internet connection;
- making clear which variables will be available in the short and long term in the datasets provided by MNOs. The availability, in the future, of also the information about gender, age, etc. of the subscriber would open the possibility of profiling the tourist, not without methodological implications due, primarily, to the imperfect match between the travelling SIM card and the SIM holder;
- ensuring the continuity and the level of territorial coverage of the data provision, taking into account also the mobile phone market shares among the providers currently operating in Italy.

The main challenge for the use of big data is to move from testing to production, considering that this evolution involves several aspects, including the need to adopt new technological and organizational infrastructures, as well as new methodological skills. With regard to this last point, the methodological approach proceeded in the experimentation could generate a virtuous circle, a continuous and iterative adjustment process with many benefits for every subject involved. On the one hand, ISTAT could achieved new methodological and technological skills to evaluate the quality and the relevance of mobile positioning data and to reduce, if it is necessary, their redundancy; on the other hand, MNOs, based on the input and feedback received by ISTAT, could improve the technique and the methodology for CDRs/DDR extraction.

Furthermore, this project represents for ISTAT a first step towards the development of a solid methodology for the use of big data, for statistical purposes, which will have positive implications not only in tourism field, but also in other areas of analysis and research, such as demographic and migration statistics.

## References

- Ahas R. (2014), *Methodological and theoretical aspects of using passive mobile positioning data for tourism marketing*, Prague – Czech Republic: 12th Global Forum on Tourism Statistics, 15-16 May 2014.
- Ahas R., Silm S., Järv, O., Saluveer E., Tiru M. (2010), *Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones*, Journal of Urban Technology, 17(1): 3-27.
- Ahas R., Aasa A., Silm S., Tiru M. (2007), *Mobile positioning data in tourism studies and monitoring: case study in Tartu, Estonia*, in Sigala, M., Mich, L., Murphy, J. (Eds.), *Springer Computer Science: Information and Communication Technologies in Tourism*, pp. 119-128.
- Bacchini, F. , D'Alò M., Falorsi S., Fasulo A., Pappalardo A. (2014), *Does Google index improve the forecast of Italian labour market?*, Cagliari – Italy: 47th Scientific Meeting of the Italian Statistical Society 2014, 11-13 June, 2014.
- Baldacci E., Scannapieco M., edit by, (2015), *ISTAT's Roadmap for the Adoption of Big Data for Official Statistics*, 20 April 2015.
- Barcaroli G., Nurra A., Salamone S., Scannapieco M., Scarnò M., Summa D. (2015), *Internet as Data Source in the Istat Survey on ICT in Enterprises*, Austrian Journal of statistics, Vol. 44, pp. 31–43.
- Barcaroli G., Nurra, A, Scarnò M., Summa D. (2014), *Use of web scraping and text mining techniques in the ISTAT survey on "Information and Communication Technology in enterprises*, Wien –Austria: European Conference on Quality in Official Statistics (Q2014) , 3-5 June 2014.
- Di Torrice M., Morricone M., Sabato M. (2012), *Beyond the CATI technique: a benchmarking analysis on demand-side surveys in EU and the Italian choice*, Reykjavík – Iceland: 11th Global Forum on Tourism Statistics, 14-16 November 2012.
- Furletti B., Gabrielli L., Garofalo G., Giannotti F., Milli, L., Nanni M., Pedreschi D., Vivi, R. (2014), *Use of mobile phone data to estimate mobility flows. Measuring urban population and inter-city mobility using big data in an integrated approach.*, Cagliari – Italy: 47th Scientific Meeting of the Italian Statistical Society 2014, 11-13 June, 2014.
- IFSTTAR, NIT, Positium LBS, Statistics Estonia, Statistics Finland, University of Tartu, (2014), *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics*, Eurostat Contract No. 30501.2012.001-2012.452, voll 1-4, <http://ec.europa.eu/eurostat/web/tourism/methodology/projects-and-studies>.
- Jonge E., Pelt M., Roos, M. (2012), *Time patterns, geospatial clustering and mobility statistics based on mobile phone network data* Statistics Netherlands, Division Methodology and Quality. The Hague/Heerlen.
- Parere sul PSN 2014-2016 - aggiornamento 2015-2016 - Registro dei provvedimenti n. 411 del 18 settembre 2014, <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3458502>.
- Scannapieco M., Virgillito A., Zardetto D. (2013), *Placing Big Data in Official Statistics: A Big Challenge?*, Brussels – Belgium: NTTS - New Techniques and Technologies for Statistics, Brussels, 5-7 March 2013
- Sonntag U., *Digging for data: Barriers to overcome before using mobile positioning data for Tourism Statistics*, Prague – Czech Republic: 12th Global Forum on Tourism Statistics, 15-16 May 2014.
- UNECE (2014), *A Suggested Framework for the Quality of Big Data. Deliverables of the UNECE Big Data Quality Task Team*.

UNECE (2014), *How big is Big Data? Exploring the role of Big Data in Official Statistics. Technical Report*, <http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=99484307>.

Vogelová M. (2012), *Using residual positioning data from mobile networks for tourism monitoring*, Reykjavík – Iceland: 11th Global Forum on Tourism Statistics, 14-16 November 2012.