



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Collection and Sources

22-24 May 2024, Geneva, Switzerland



Mobile Phone Data for Enhanced Tourism Statistics in Italy: Insights from Vodafone-Istat Project Foundation

Maria Teresa Santoro, Lorenzo Cavallo and Silvia Di Sante (Istat, Italy)

LORENZO CAVALLO

Istat | Directorate for Environmental and Territorial Statistics (DCAT) - Rome

Introduction

- This work provides an overview of the activities and findings from the concluding phase of the **Project Foundation**, a collaborative initiative between Vodafone (a Mobile Network Operator) and Istat to share this national experience from a thematic point of view
- Started with a preliminary phase, called Sprint, conducted in Sept-Nov 2020, and aimed at a first exploration of the potential of mobile data in provide estimations/trends on arrivals and nights spent in Italy (*domestic tourism* and *inbound tourism*), the PF is exploring some aspects in depth, first of all: **the convergence improvement of the conceptual definitions and classifications**
- Some of the **themes** investigated:
 - possibility to use mobile data for further disaggregate tourist flow sat a territorial and/or temporal level
 - possibility to exploit its timeliness
 - possible use of mobile phone data to do estimates, forecasts or calibrations of surveys samples
 - possible new themes of interest

Overview of official statistics on tourism

The information on tourism are fundamentals and increasingly requests from many types of users of data, in particular *Policy makers*, national and locals, Operators of the sector, Researchers and analysts

Turning to alternative data sources to produce official statistics has become an increasingly focal theme.

What happened around us?

- ✓ Evolution of digital technologies growth: platforms of booking *online*, sites of travel reviews, social network, Telco operators,...) are all **non-traditional sources of data** with possible impact on world of official statistics and, in particular, tourism data sector
- ✓ Vast amount of information (*Big data*), collected for non-statistical purposes but potentially useful for strengthening the information framework on tourism

There is growing interest in these data

- ✓ Final objective: evaluate the possibility of use to improve or produce (new?!) statistics relevant and reliable
- ✓ A unique challenge for the world of official statistics

Use of Mobile Network Operator data for tourism statistics



- ✓ Some experiments also at the level international and national
- ✓ At first glance, very promising for complementing and enriching current investigations on the tourism in particular for the ability to provide information very **timely** and with high **territorial detail**

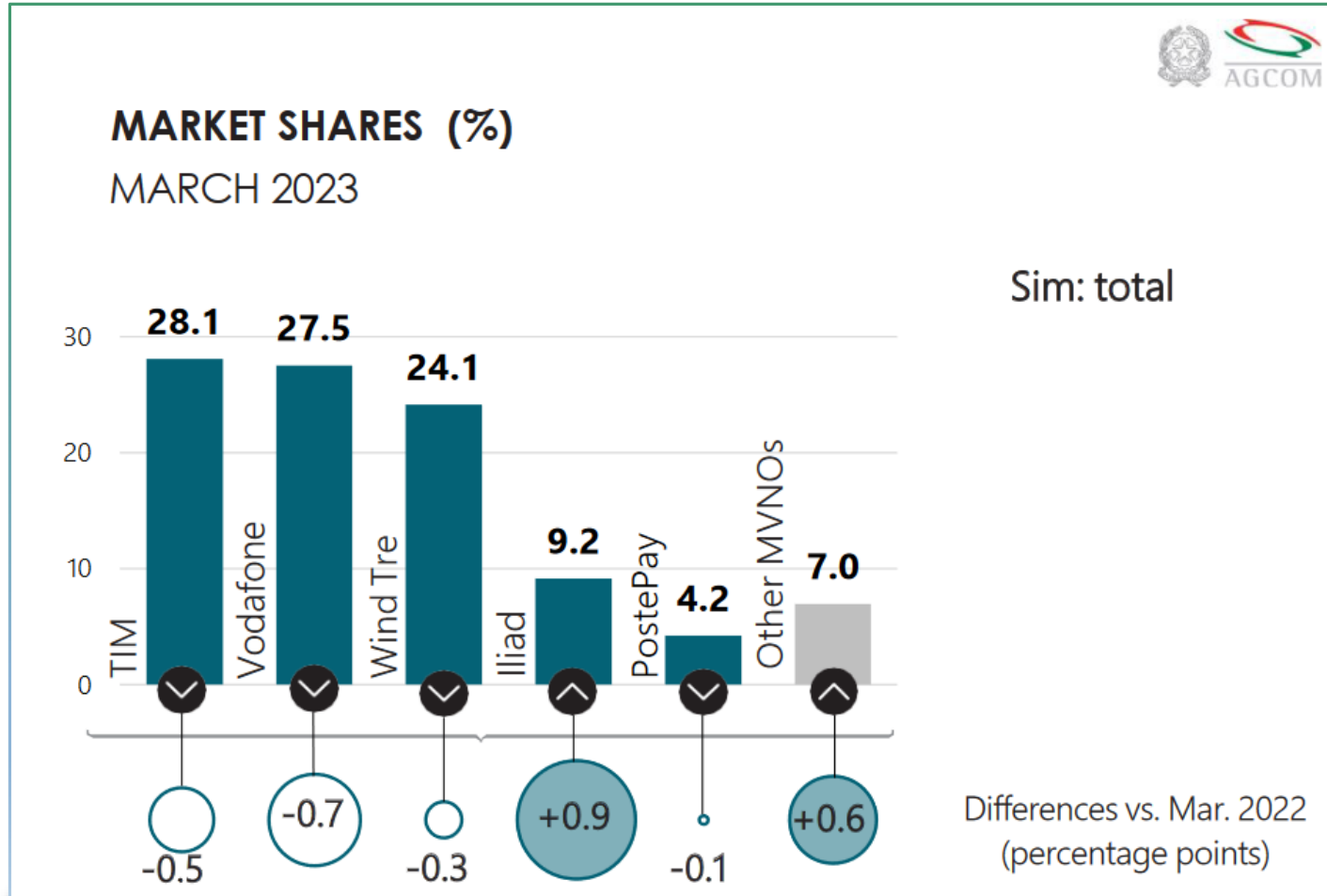


- **First collaboration protocol** between Istat and Vodafone Business Italy (2019) (also with Motion Analytica) for the identification of methodologies innovative statistics and of quality (not for exchange of data!) and test the potential use of mobile phone data in the production of official statistics (including the **tourism** sector)

Mobile Network Operator



Features of Vodafone



- ❖ Beyond 200,000 cells telephone calls
- ❖ Covers almost 99% of the Italian territory
- ❖ The data are collected every minute
- ❖ The spatial data resolution is greater in cities and less in rural areas

Mobile Network Operator



Characteristics of data

We can say that the MNO data provide information on:

- **Who** → Code SIM
- **Where** → Area antenna coverage
- **When** → Day/Time

What DON'T they tell us?

**WHY the SIM is in an area
And WHY is moved there**

- We accept some forcing on the WHY (one SIM \neq one person) and WHERE (territorial assignment based on cells may not be precise)
- MNO data they are defined as personal data and anonymized under privacy protection

Technology used by Vodafone Analytics: intercepts and collects in real time the continuous flow of data between *smartphones* and the mobile phone network (*signaling data*), ensuring temporal and spatial information granular, through proprietary (and confidential) estimation algorithms

Cell tower records allow for generating a table that identifies, minute by minute, the connecting cell for each user. Each cell is associated, based on its coverage area, with one and only one municipality; therefore, in each minute, each user is assigned to a specific municipality.

Methodological Issues

- ❑ One of the main challenges encountered is how to relate the **resident population** and the **Telco users** (i.e. all people with *smartphones/SIM*), to do this several problems must be addressed:
 1. how to make the subscribers of a single operator representative of the population resident
 2. how to consider people who do not have or do not use *smartphones/SIM*
 3. how to treat those who sign more than one contract
- ❑ Further problems are linked to the identification of **places** (i.e. place of residence or place of tourism) based on telephone locations
- ❑ All information provided is extended to the entire reference universe
- ❑ A **proprietary algorithm** allows to represent the entire universe of users and not just Vodafone SIM owners or foreigners connected to the Italian network. It is based on:
 1. Local market share of the operator (from internal market analysis of Italian SIMs by province and age groups)
 2. National market share of foreign SIMs
 3. Market share by type of SIM from market studies and official reports (AGCOM)
 4. Socio-demographic characteristics of users from proprietary data

From *Sprint* to *Project Foundation*

Sprint on tourism

- ✓ **Perimeter** of the analysis: Province of Rimini and the Municipality of Rome (2 areas with a strong tourist vocation)
- ✓ **Time periods**: 3 periods August 2019 and 2020 and April 2020 (period of *lockdown* with tourist flows equal to 0)
- ✓ **Case study** (from **supply side**): arrivals and nights spent of tourists resident in Italy (*domestic*) and coming from abroad (*inbound*)



Project Foundation

- ✓ **Perimeter** of the analysis: Emilia Romagna and Lazio Regions (*detailed data at municipal, provincial and regional level*)
- ✓ **Time periods**: From June 2022 to October 2023
- ✓ **Case study**: arrivals and nights spent by residence of the guests (**Regions** for domestic tourists, **Countries of residence** for international tourists)

Definitions finalized during the *Sprint*

Concepts	Definitions officials	Definitions adopted
Place of usual residence	Municipality of usual residence of the tourist	<p>Prevalent Night Cell: For the attribution of the prevalent night cell, the twelve hours between 8 PM and 8 AM the following morning are considered.</p> <p>Prevalent municipality: It is the municipality where the user spends the most hours. To define a valid overnight stay, the user must spend at least six hours within the municipality.</p> <p>Telephone residence: corresponds to the municipality of the prevailing night cell, in the last 12 months before the reference period</p>
Tourist	Traveler making an overnight stay in a destination outside the municipality of usual residence, for any purpose	A user in a prevalent night cell in a municipality other than the telephone residence (defined, therefore, as municipality of destination)
Arrivals	Number of clients who carry out the <i>check-in</i> in accommodation establishments (of a given municipality) in the reference period	A «tourist» is registered in the municipality of destination as an «arrival» on the first day of his trip
Nights spent	Number of nights spent by customers in accommodation establishments (in a given municipality) in the reference period	Sum of nights spent by «tourists» in the reference period, where «nights» = daily sum of «tourists» in the municipality of destination
Resident tourist (domestic)	Tourist resident in Italy	Tourist with an Italian SIM or a Foreign SIM (in this case if present in Italy for at least 6 months, therefore defined as a Foreign resident user)
Non resident tourist (inbound)	Tourist Non resident in Italy	Tourist with Foreign SIM with the exclusion of Foreign resident users

Improvements introduced with the *Project Foundation*

The *Project Foundation* allowed us to make some steps forward in some **Definitions**

TRAVELLER CLASS

This attribute represents a macro-distinction of users based on their residence, which is attributed based on the Mobile Country Code (MCC) of the SIM.

A user is thus classified as:

«**Resident Users**»: users who use an Italian SIM or users who use a foreign SIM but are present in Italy for at least 90 days in the 12 reference months

«**Non resident Users**»: users who do not use an Italian SIM and who have spent less than 90 days in Italy in the 12 reference months

USUAL RESIDENCE

The usual residence of a user is the prevalent night time municipality, i.e., the municipality where the user has spent the most nights in the twelve reference months and is defined only for Italian users.

Usual Environment

The **Usual Environment** of a user is **the set of municipalities** where the user is habitual; **their presence in these municipalities does not constitute tourist behavior**

Only for Resident users

Given a number of months and a total number of hours that the user has been present in a specific municipality, that municipality is considered to belong to their **usual environment** if it exceeds the hourly threshold for that particular number of months

More inclusive definitions of the usual environment correspond to a contraction of the tourist phenomenon because they increase the set of municipalities where the user is not considered a tourist

Various definitions of UE have been experimented with, and for each of these definitions

For each of these definitions, a version with a «buffer» was also tested, which adds to the set of municipalities obtained by applying that particular definition all the neighboring municipalities of a municipality in that set

Main Results

- ✓ The ***Project Foundation*** undoubtedly allowed us to make gods **steps forward** compared to ***Sprint***
- ✓ The study on **Usual Environment** allowed us to improve the estimates of domestic tourism
- ✓ Attendance calculated with *Telco* data are almost always greater than those detected by Istat (*Telco data* do not allow us to distinguish where the tourists stayed)
- ✓ The estimate of arrivals seems improved compared to *Sprint* but it would still be necessary do insights and perfect the definitional convergence
- ✓ It should be underlined that this evidence relates to only the two territories analysed: a national analysis could lead to different results and therefore caution must be used

Evaluations and prospects

- ✓ To the moment we are still far away from the possible use of mobile phone data *tout court* to produce official statistics on the topic of tourism. However, the analyzed trends have highlighted that it might be possible to use these data to estimate territorial trends
- ✓ It is also necessary to further work on refining the definitional convergence
- ✓ The MNO data can open new analytical perspectives in the tourism sector, exploiting territorial granularity and the eldest timeliness
- ✓ The experience has confirmed the potential of the data MNO
- ✓ It is difficult to imagine that mobile data can replace traditional data sources for the production of statistics on tourism
- ✓ In brief term the MNO data are not ready for use: interaction is necessary constant with the mobile operators to test definitions, algorithms and evaluate the quality of the data

Thank you

LORENZO CAVALLO | cavallo@istat.it