METHODOLOGICAL ASPECTS FOR PASSENGER MOBILITY ANALYSIS USING BIG DATA

MARIO TARTAGLIA
Leader Of FS Research Centre Ferrovie dello Stato Italiane

GIOVANNA ASTORI
ISTAT

10th ITF Statistics Meeting OECD Conference Centre Paris, 11-12 April 2024
Contents

❖ Why a «Technical paper» on Big Data for passenger mobility analysis?
❖ Two questions
❖ The legal frame for passenger mobility statistics
❖ Available statistical and administrative sources on passenger mobility
❖ Big Data for passenger mobility statistics: promising sources
❖ Processing Big Data: the life cycle pipelines. Critical issues
❖ Case studies: MPD and other BD sources for Passenger mobility analysis
❖ The way forward
Why a «Technical paper» on Big Data for PM analysis?

✓ From life to «onlife»: people in their daily life generate huge amounts of signals continuously.

✓ “[onlife is a neologism coined in order to refer to] the new experience of a hyperconnected reality within which it is no longer sensible to ask whether one may be online or offline” (The Onlife Manifesto, L. Floridi. 2015)

✓ The «6 V» of Big Data: Volume, Velocity, Variety, Veracity, Value, Variability

✓ Big Data and Small/Thick data: two legs that walk together

✓ Are Big Data useful to represent passenger mobility?

✓ Let’s start building a map to dive into the ‘data waves’!

✓ Priority project in the Trusted Smart Statistics Road map by Istat (Italian NSI)
Two questions

1. Which perspective are we interested in?
2. How to measure and describe mobility?
The legal frame for passenger mobility statistics

❖ Eurostat Guidelines on Passenger Mobility statistics (2018): methodological frame (NO Regulation)

❖ EU Regulations on statistics by mode of transport (Air, Sea, Rail, IWW. NO Road – private car)

❖ Italian laws for the monitoring of mobility supply/demand (SUMPs – Sustainable Urban mobility planning and monitoring. Mobility Manager monitoring of ‘home to work/school’ trips)
Available statistical and administrative sources on passenger mobility

❖ Statistics on mobility behaviour

✓ **AUDIMOB** by Isfort – household travel survey, compliance with Estat PM Guidelines

✓ Istat (italian NSI) surveys not aimed at describing PM, that collect data on mobility also: Population census, HETUS, Multiscope - Aspects of daily life etc. (limited on some aspects)

✓ Local administrative units Travel surveys (limited in space)

❖ Statistics on supply/demand of transport and on infrastructures

❖ Statistics by mode of transport and Traffic statistics (Vkm)
Big Data for passenger mobility statistics: promising sources

- Mobile Phone Data
- Automatic Identification System data
- Social network data
- Sharing mobility data
- Land use data
- Smart cards data
- Floating car data
- Other IOT data
- GPS tracking
- Points of interest data (POI)
- Transport industry data
- GPS tracking
- Other IOT data

Sources:
- Points of interest data (POI)
- Floating car data
- Smart cards data
- Transport industry data
- IOT data
- Land use data
- Social network data
- Sharing mobility data
- Automatic Identification System data
- Mobile Phone Data
Due to their nature, Big Data need specific procedures to be processed.

A dedicated pipeline must be drawn in order to consider all the steps that lead from the identification of the new data source to its final use as a piece of information for statistical production (Big Data Life Cycle).

Various proposals have been implemented for the handling and treatment of Big Data.

In particular, the focus is on BREAL – Big data REference Architecture and Layers (Eurostat).

Approaches to privacy preserving.

(adapted from ESSNET Big Data II – WPF – Deliverable F1, 2019)
In the BD lifecycle models, new figures and skills emerge with respect to the data treatment process.

The BREAL model identifies some categories of ‘actors’ that contribute to the process.

- IT & Statistical pipeline actors
- Capacity providers
- Audit, control and compliancy actors
- Global roles: citizens, statistical institutions, system orchestrator, medias, government

(adapted from ESSNET Big Data II – WPF – Deliverable F1, 2019)
## Processing Big Data: critical issues

<table>
<thead>
<tr>
<th>General issues</th>
<th>Methodological/thematic issues</th>
<th>Data processing issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Accessibility</td>
<td>✓ Relation between the infrastructure that generate BD and the thematic dominions</td>
<td>✓ IT architecture</td>
</tr>
<tr>
<td>✓ Interoperability</td>
<td>✓ Reference population</td>
<td>✓ IT infrastructure</td>
</tr>
<tr>
<td>✓ Security, protection, privacy preserving</td>
<td>✓ Significance of data as information</td>
<td>✓ Process models and pipelines</td>
</tr>
<tr>
<td>✓ Privately held data</td>
<td>✓ Bias</td>
<td>✓ Algorithms</td>
</tr>
<tr>
<td>✓ Data Act</td>
<td>✓ Sample coverage</td>
<td>✓ Privacy preserving</td>
</tr>
<tr>
<td>✓ Open data</td>
<td>✓ ...</td>
<td>✓ AI Act</td>
</tr>
<tr>
<td>✓ Trust</td>
<td>✓ ...</td>
<td>✓ Training, skills</td>
</tr>
<tr>
<td>✓ ...</td>
<td>✓ ...</td>
<td>✓ Team</td>
</tr>
</tbody>
</table>

April 2024

Tarteglia M, Astori G | Methodological Aspects for Passenger mobility Analysis using Big data
Case studies: MPD and other BD sources for PM analysis

Case study: a nationwide mobility study (Italy)

- single telephone service provider with about 23 million Human-type SIM cards.
- Definition of the “journey”: concatenation of all possible intermediate stops lasting less than one hour within the context a single origin-destination displacement.
- Zoning system: the nationwide study area was broken down using the municipalities boundaries.
Case studies: MPD and determining the mode of transport

Mode of transport: identifying the rail journeys

- associating a set of reference radio base stations with each of the monitored railway stations
- identifying a series of points called “railway gates”, points along the rail network whose mobile network coverage does not reach the main roadways

Validation:

- The ideal train user is not seen passing through any “road gates” (points at which trains do not transit)
- The “theoretical” sequence of stations (using national rail network)
- Set of intermediate stations for each origin/destination pairing.
Case studies: MPD and dissemination

https://www.fsitaliane.it/content/fsitaliane/en/fs-research-centre/mobility-in-italy.html

https://www.fsitaliane.it/content/fsitaliane/en/fs-research-centre/station-usage-data.html
The way forward

- More case studies
- Guidelines
- International projects
- ...

The technical paper (Italian and English version) is available for free at:
https://issuu.com/ferroviedellostatoitaliane/docs/fsrc_eng_technical_papers_2024
Thank you!

MARIO TARTAGLIA
GIOVANNA ASTORI