

TECHNICAL PAPERS

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



MINISTERO
DELLE INFRASTRUTTURE
E DEI TRASPORTI



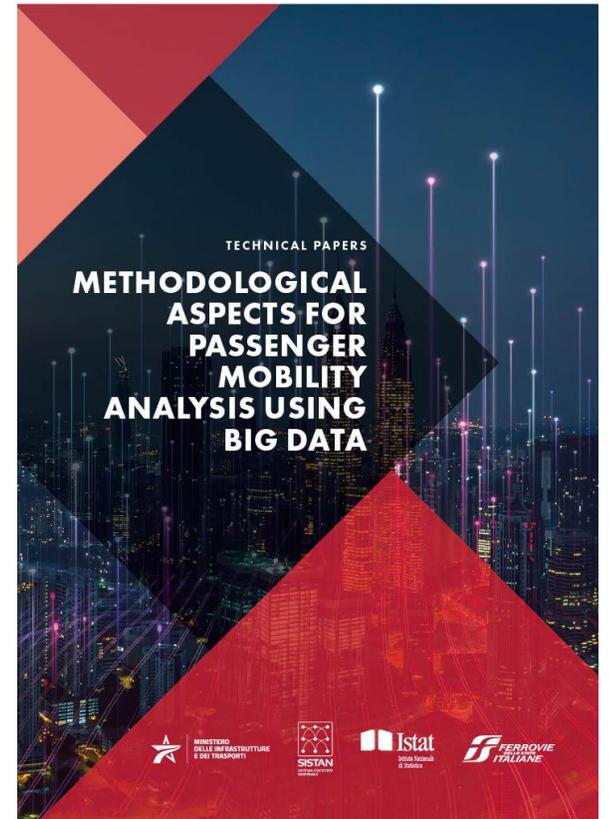
SISTAN
SISTEMA STATISTICO
NAZIONALE

Istat
Istituto Nazionale
di Statistica

FS FERROVIE
DELLO STATO
ITALIANE

Contents

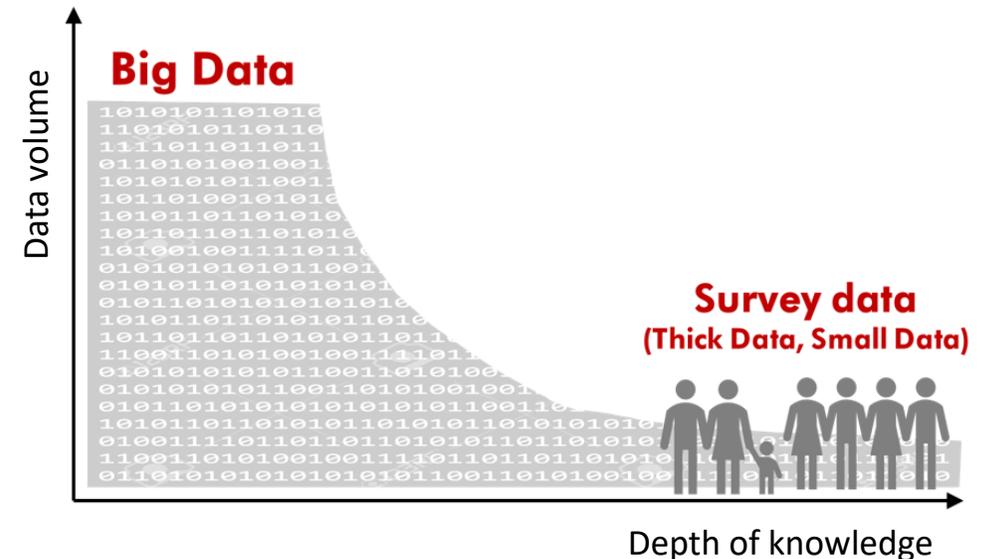
- ❖ Why a «Technical paper» on Big Data for passenger mobility analysis?
- ❖ Two questions
- ❖ The legal frame for passenger mobility statistics
- ❖ Available statistical and administrative sources on passenger mobility
- ❖ Big Data for passenger mobility statistics: promising sources
- ❖ Processing 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)

The relationship between Big Data and traditional surveys

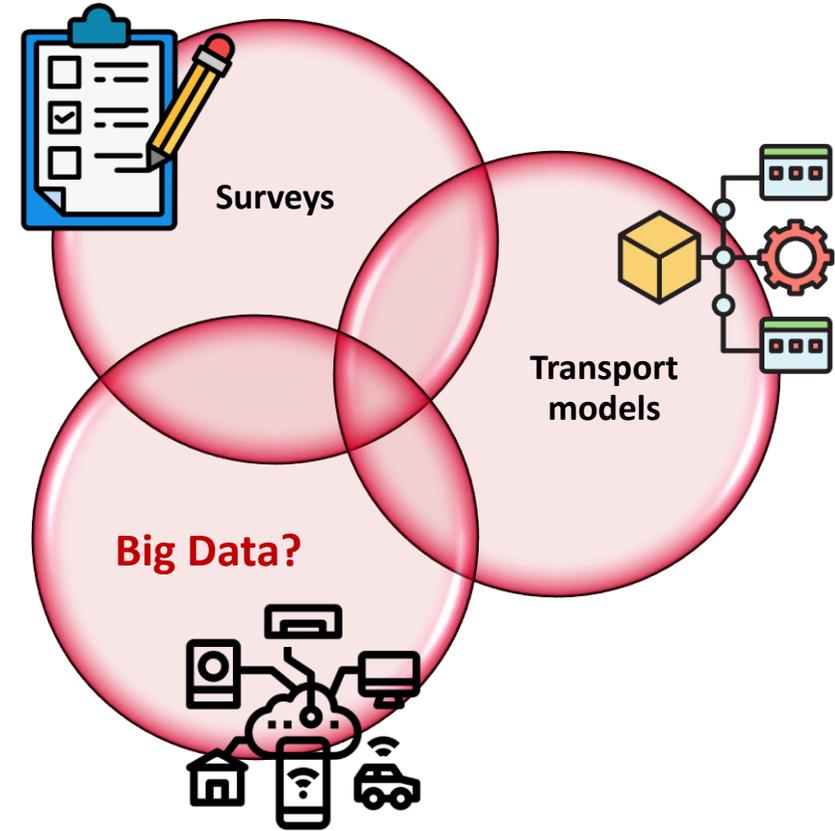


(freely adapted from Tricia Wang, 2016)

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 (SUMP – Sustainable Urban mobility planning and monitoring. Mobility Manager** monitoring of ‘home to work/school’ trips)

Available statistical and administrative sources on passenger mobility

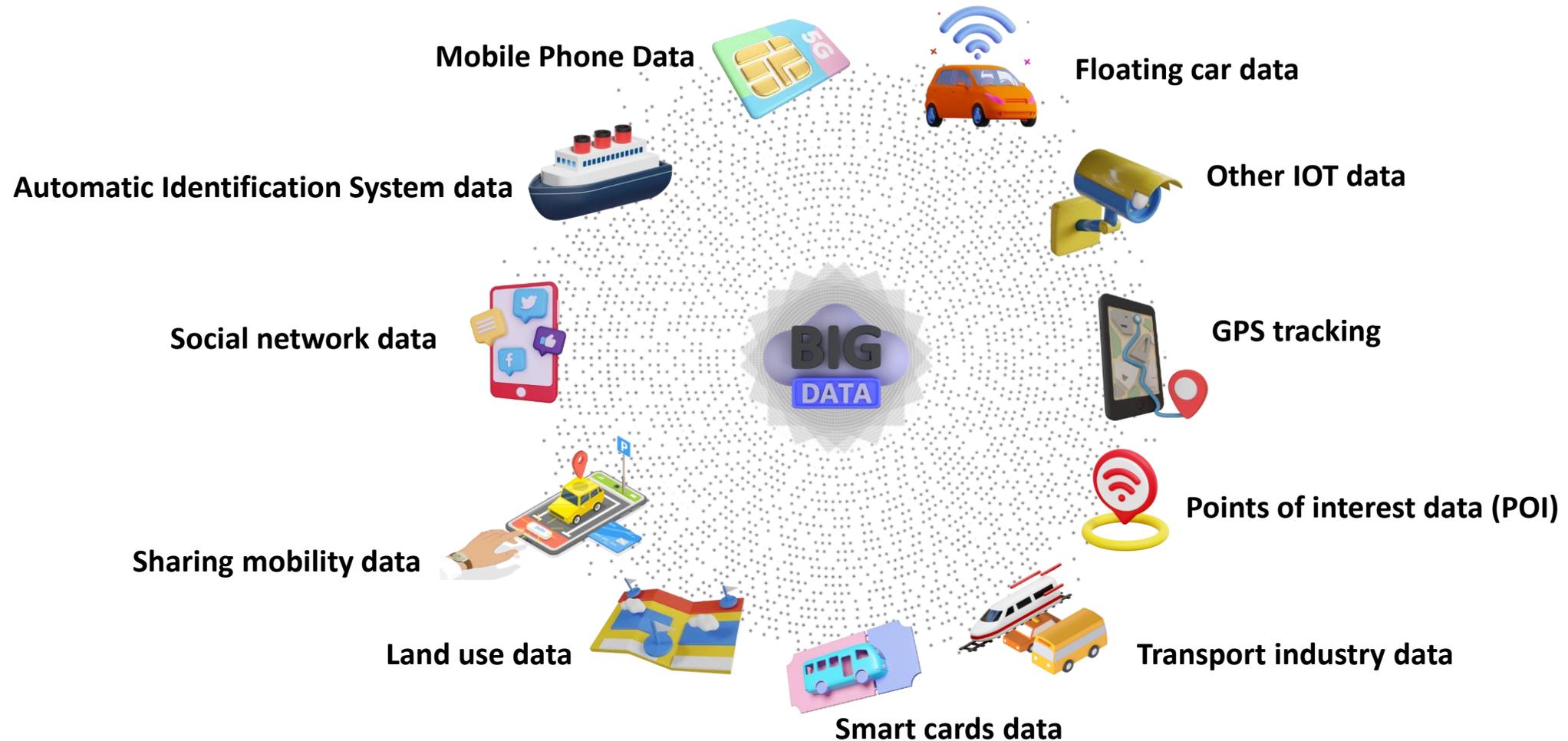
❖ Statistics on mobility behaviour

- ✓ **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)

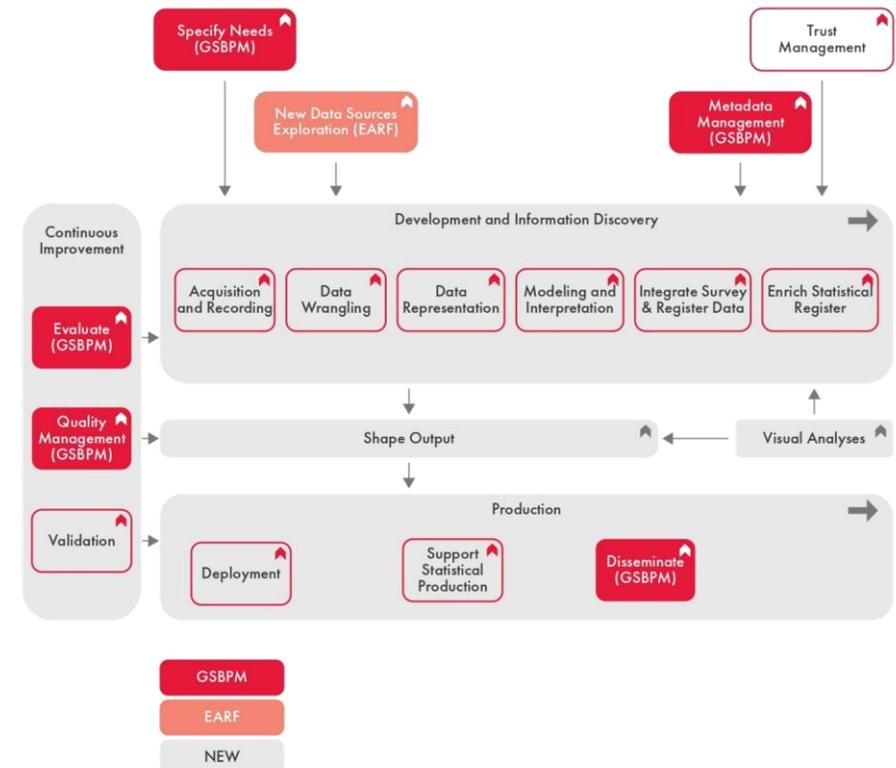
Big Data for passenger mobility statistics: promising sources



Processing Big Data: the life cycle pipelines

- 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

BREAL Big Data Life Cycle

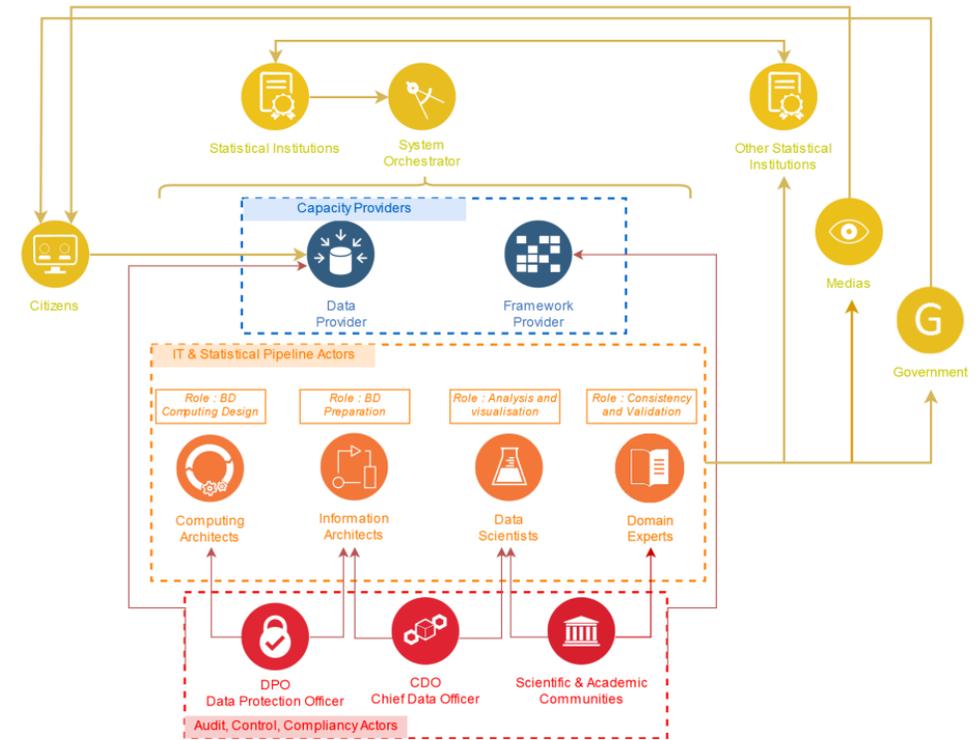


(adapted from ESSNET Big Data II – WPF – Deliverable F1, 2019)

Processing Big Data: the life cycle pipelines

- 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

BREAL model Actors



(adapted from ESSNET Big Data II – WPF – Deliverable F1, 2019)

Processing Big Data: critical issues

General issues

- ✓ Accessibility
- ✓ Interoperability
- ✓ Security, protection, privacy preserving
- ✓ Privately held data
- ✓ Data Act
- ✓ Open data
- ✓ Trust
- ✓ ...

Methodological/thematic issues

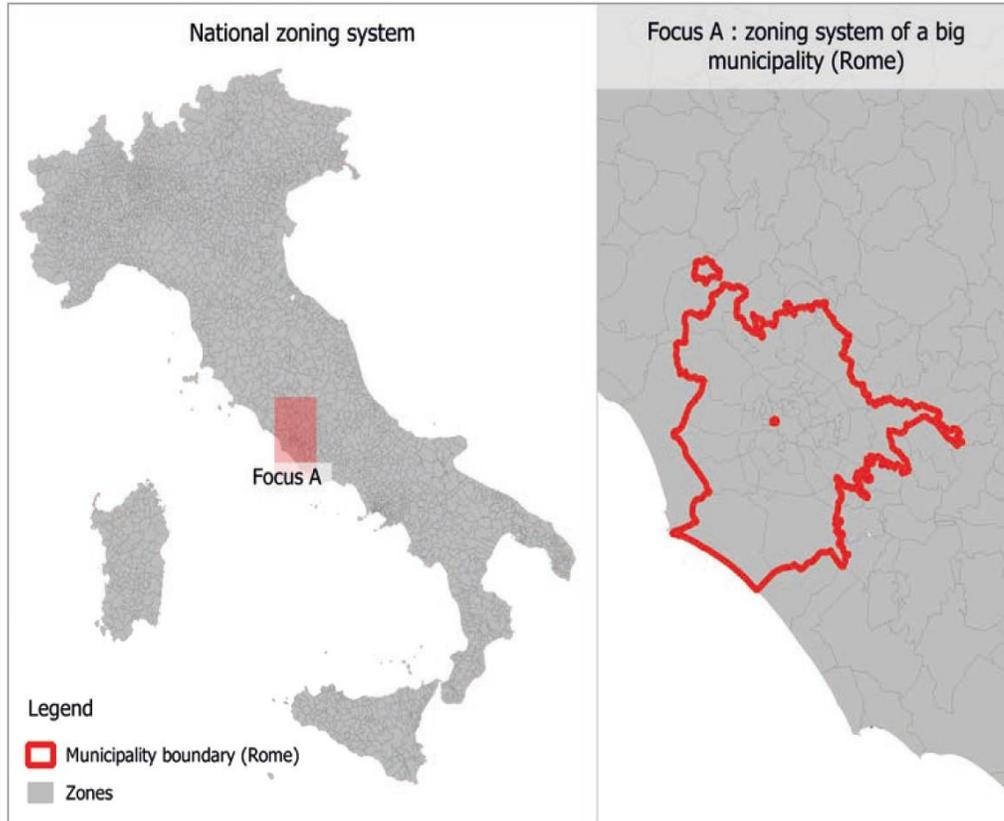
- ✓ Relation between the infrastructure that generate BD and the thematic dominions
- ✓ Reference population
- ✓ Significance of data as information
- ✓ Bias
- ✓ Sample coverage
- ✓ ...

Data processing issues

- ✓ IT architecture
- ✓ IT infrastructure
- ✓ Process models and pipelines
- ✓ Algorithms
- ✓ Privacy preserving
- ✓ AI Act
- ✓ Training, skills
- ✓ Team
- ✓ ...

Case studies: MPD and other BD sources for PM analysis

Case study: a nationwide mobility study (Italy)

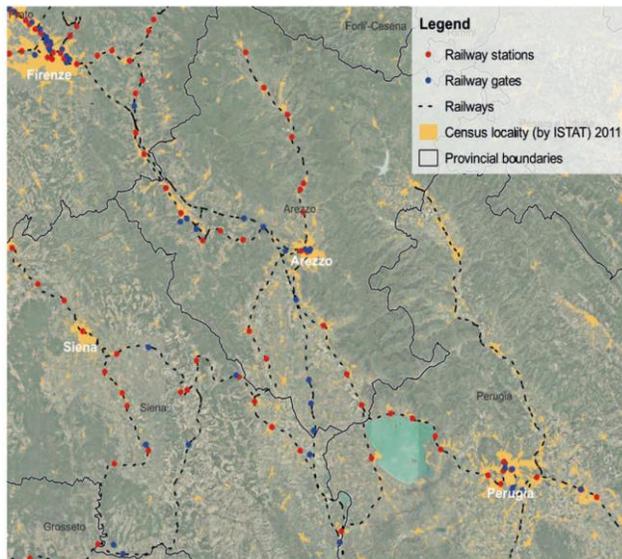


- ✓ single telephone service provider with about 23 million Human-type SIM cards.
- ✓ Definition of the “journey”: concatenation of all possible intermediate stops lasting less than one hour within the context a single origin-destination displacement.
- ✓ Zoning system: the nationwide study area was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

Case studies: MPD and determining the mode of transport

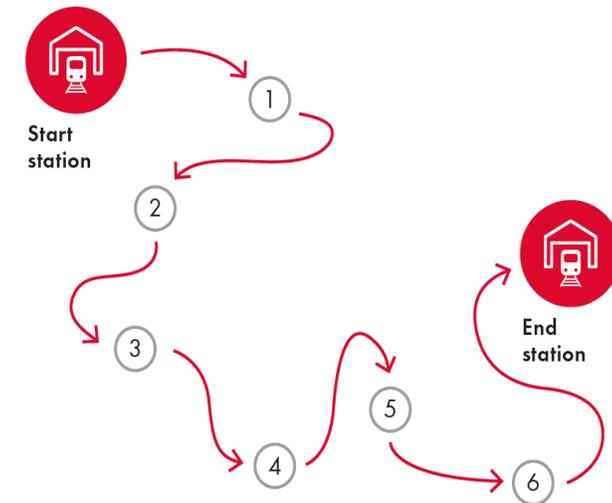
Mode of transport: identifying the rail journeys

- 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

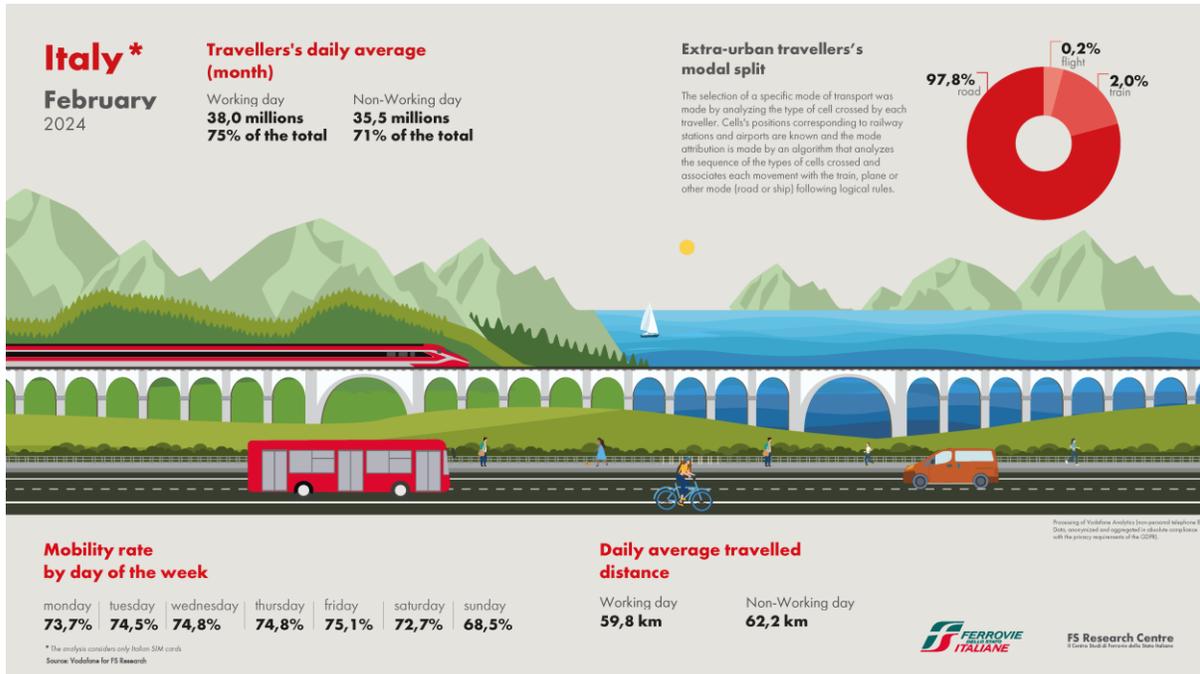


Validation:

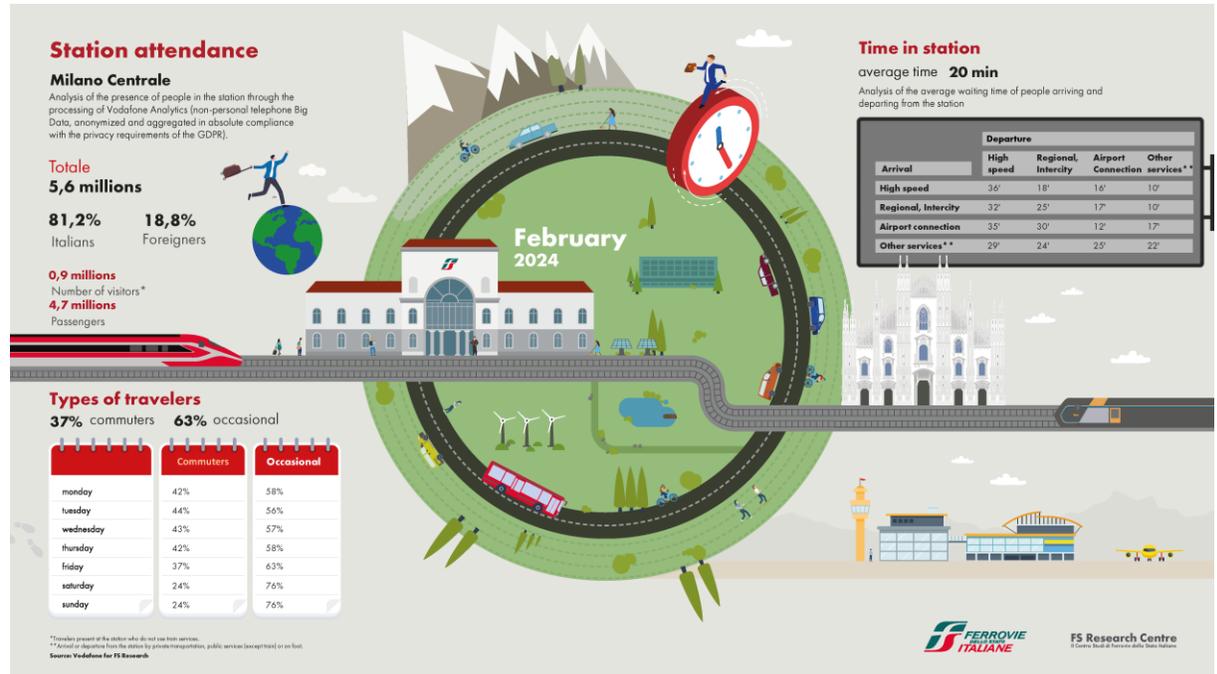
- The ideal train user is not seen passing through any “road gates” (points at which trains do not transit)
- The “theoretical” sequence of stations (using national rail network)
- Set of intermediate stations for each origin/destination pairing.



Case studies: MPD and dissemination



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



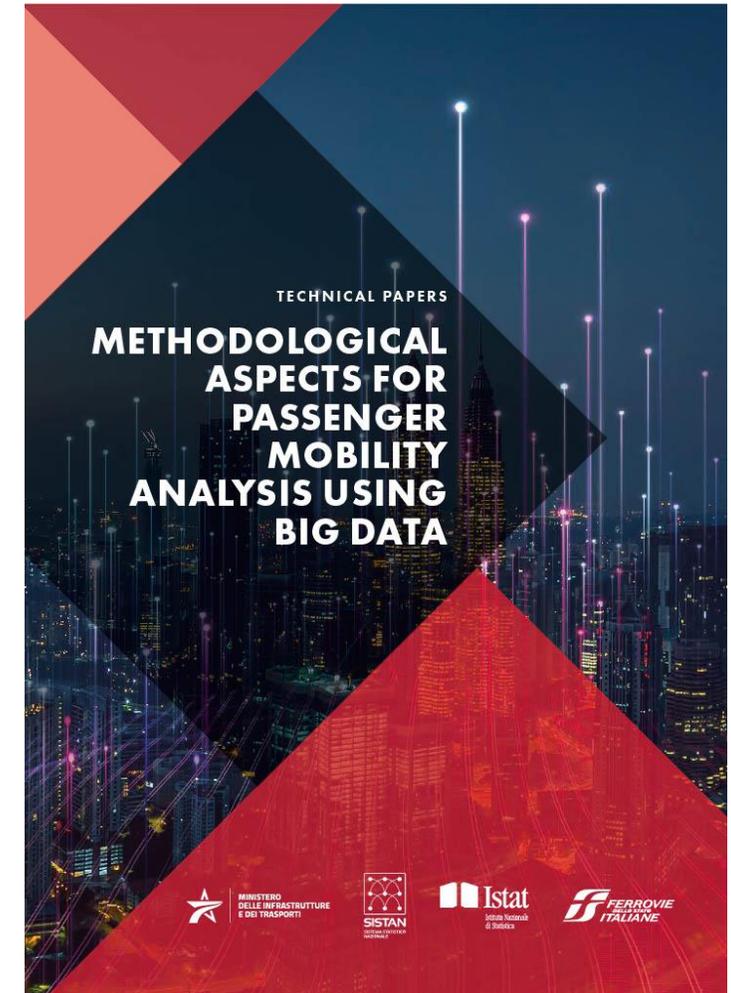
<https://www.fsitaliane.it/content/fsitaliane/en/fs-research-centre/station-usage-data.html>

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



A night cityscape with a digital network overlay. The background shows a city at night with illuminated buildings and streets. Overlaid on this is a complex network of glowing lines and dots in various colors (blue, purple, pink, green) that represent a digital or data network. The lines connect various points across the city, creating a sense of connectivity and data flow.

Thank you!

MARIO TARTAGLIA
GIOVANNA ASTORI