July 2025

F.A.Q. English

USE OF OPEN STREET MAP FOR ACCIDENT INVESTIGATION ON THE ROAD AND MOTORWAYS NETWORKS. Updating year 2023.

A) General Concepts and Project Objectives

1.   What is the main purpose of this project?

The project aims to calculate road crash, fatality, and injury indicators using road network length as the denominator to provide a more accurate measure of risk exposure.

2.   What is the ultimate long-term goal?

The ultimate goal is to estimate actual traffic flows (vehicles/km) on the national road network to calculate the real probabilities of being involved in a crash and the true rates of risk exposure.

3.   What are Road Safety Performance Indicators (RSPIs)?

They are multi-dimensional indicators that analyze crashes, roads, vehicles, and people to support planning, reduce fatalities and injuries, and evaluate the effectiveness of road safety policies.

4.   Why traditional denominators like "resident population" are problematic?

The resident population does not adequately reflect user mobility, ignoring the seasonality of crashes and the concentration of traffic in specific locations, which can lead to distorted indicators.

5.   What are the limitations of the "vehicle fleet" denominator?

The vehicle fleet by province of registration does not account for the actual mobility of users on the territory. A vehicle may be predominantly driven in another province than the one where it is registered.

6.   What is the main innovation this project introduces to statistical production?

The innovation lies in the integration of different sources (official statistics, GIS, Big Data like OSM), enhancing existing administrative sources, and introducing new methodologies to improve the quality of statistical information.

7.   Is the project considered an official Istat statistic?

The project has been included among the priority projects of the Trusted Smart Statistics (TSS), which involve the use of Big Data and a rigorous methodological evaluation. It may become an official statistic after meeting the necessary requirements.

8.   What data was used for this update?

This update uses road crash data from the year 2023.

9.   Why was a long time interval needed for this update?

To maintain high-quality standards, it was necessary to wait for the release (which occurred in 2023-2024) of the new Istat territorial bases and Census tracts updated to 2021, which are essential for the road reclassification methodology.

10.   How do the results help road safety stakeholders?

They provide nationally standardized tools for better network management, crash analysis, and the identification of improvement that is more effective and prevention actions.

B) Data Sources and Tools (GIS, OSM, Istat)

1.   What are the main data sources used?

The main sources are the Istat survey on road crashes, geographic data from OpenStreetMap (OSM), and Istat's territorial bases.

2.   What is OpenStreetMap (OSM) and why is it used?

OSM is a collaborative project to create a free, editable map of the world. It is used to obtain detailed data on the road network (e.g., length per direction of travel, road type), information not available in a unified national road registry.

3.   Is OSM a reliable data source for experimental statistics?

Yes, OSM is considered reliable and robust, so much so that it is used by most GPS navigation software. Furthermore, the community updates the data daily.

4.   What is a Geographic Information System (GIS)?

It is a system that allows for the capture, analysis, management, and graphical representation of georeferenced (spatial) data.

5.   What is the key role of GIS in this project?

GIS allows for the correlation of information from different sources (crashes, OSM road network, Istat administrative boundaries) using geographic location as the linking key.

6.   What are "map layers" and how are they used?

They are overlaid levels of geographic information. In the project, OSM layers (roads) and Istat's bases (localities) are overlaid to assign the correct context (e.g., urban or non-urban) to each road.

7.   What are "Istat's territorial bases"?

They are the official geographic data defining the boundaries of Italian administrative units: Regions, Provinces, Municipalities, and Localities.

8.   How are "localities" classified by Istat?

They are classified into four categories: Urban areas, Inhabited centers, Production sites, and Scattered houses.

9.   What types of OSM roads are included in the analysis?

All road types intended for motor vehicle circulation are included, such as motorways, national, provincial, municipal, and service roads. Cycleways, footpaths, etc., are excluded.

10.   Is there an official and harmonized national road registry?

No. Although local-level archives and road graphs exist, a systematic national road registry has not yet been established.

C) Calculation and Classification Methodology

**1.	How is road length calculated in an innovative way?**
The calculation is not limited to the length of the road segment but also considers the number of carriageways for each direction of travel, providing a more precise measure of the available infrastructure.

**2.	What is a "spatial join" operation?**
A GIS operation merges data from different layers based on their geographic location. For example, it allows a road segment (from OSM) to be associated with the locality (from Istat) in which it is located.

**3.	What is meant by a "systematic" classification of roads?**
It is a methodology that links OSM's functional categories (e.g., Primary, Secondary) to Istat's management-related ones (e.g., Urban Road, Non-urban Road) through a linking table based on location (urban/non-urban).

**4.	What is meant by an "analytical" classification, and why is it an evolution?**
It is a more advanced classification that is not just based on fixed rules but analyzes the metadata of each individual road segment (name, reference code, OSM class, and locality). It allows for overcoming the limits of the systematic approach and grouping roads in a more detailed and accurate way according to their management characteristics.

**5.	What are the advantages of the analytical classification?**
It allows for better management of the road network, more effective crash analysis, and the construction of more precise and standardized risk exposure indicators.

**6.	What are Traffic Points (PoT) and where do they come from?**
They are points, extracted from OSM, that locate a significant traffic intensity on a road segment. Over 390,000 were counted in 2023.

**7.	How are PoTs used in calculating the indicators?**
They are used to "weight" the indicators. The length of roads with traffic points is calculated to "adjust" the indicator for the effect of traffic volume, thereby isolating other risk factors such as infrastructure weaknesses or driver behavior.

**8.	What does it mean to "adjust the indices for the traffic factor"?**
It means calculating the crash risk net of traffic intensity. This highlights the intrinsic danger of a road segment, regardless of how many vehicles use it.

**9.	What effect does this "adjustment" have on risk assessment?**
For the same number of crashes, a road segment with less traffic will appear relatively more dangerous, highlighting problems not related solely to congestion.

**10.	How is the correspondence between different road classifications (OSM and Istat) managed?**
Through the creation of linking tables, which associate OSM's functional categories with those used by Istat in its crash survey, also taking into account the geographic location.

D) Main Results and Analysis

**1.	What are the three main sets of indicators calculated?**
Crash indicators are calculated relative to: 1) road network length, 2) vehicle fleet, and 3) resident population.

2.   What does the indicator based on road network length show?
It measures the number of crashes, fatalities, and injuries per 100 kilometers of carriageway, providing a measure of risk directly linked to the infrastructure.

3.   Which areas show the highest risk according to the indicator based on road length?
Large urban centers (e.g., Milan, Genoa, and Rome) show the highest rates for crashes and injuries on motorways and urban roads. Medium-sized provinces (e.g., Monza and Brianza) stand out for non-urban roads.

4.   What does the correlation matrix between the different indicators show?
It shows that indicators based on population and vehicle fleet are highly correlated with each other (0.88) but less correlated with the indicator based on road network length. This confirms that measuring risk based on infrastructure captures a different dimension of the phenomenon.

5.   What is the most important result of the analysis with Traffic Points (PoT)?
The analysis shows that the risk ranking of some provinces changes when the indicator is "adjusted" for the traffic factor. For instance, the danger of some roads in Bologna or Rome increases relative to those in Prato or Venice, highlighting a greater infrastructural or behavioral risk.

6.   What percentage of the Italian road network has Traffic Points (PoT)?
Approximately 3% of the total network, corresponding to over 75,000 km of carriageway.

7.   Which road types have the highest concentration of PoTs?
Trunk roads (14.1%), primary roads (9.0%), and secondary roads (4.2%).

8.   Does using the analytical classification significantly alter the results compared to the systematic one?
No, the correlation between the two methods is very high (0.99). However, the analytical method represents a qualitative improvement that makes the measurement more objective and aligned with the actual management of the roads.

9.   Which provinces rank highest for overall risk according to the composite index based on road length?
The provinces of Milan, Monza and Brianza, Rome, Naples, and Florence.

10.   Why can indicators based on population produce "unrealistic" values on motorways?
Because the resident population of a province does not include the huge flows of non-resident users (commuters, tourists) who travel through major motorway hubs, thus distorting the risk rate.

E) Composite Indices and Ranking Methodology

1.   What is a "composite index" in this context?
A single value mathematically combines several elementary indicators (e.g., crash, fatality, injury rates) to measure a complex concept like the "road crash risk" of a province.

2.   What software tool was used to build the composite indices?
The standard Istat tools "RankerTool" and "i.Ranker" were used, which allow for the calculation, visualization, and comparison of composite indices using different methods.

3.   Which aggregation method was chosen and why?
The Mean of z-scores (MZ) method was chosen because it is well known, easy for all users to interpret, and showed good robustness in tests conducted with the COMIC software.

4.   What are the fundamental steps to build a composite index with Ranker?
The two main steps are: 1) the standardization of the elementary indicators to make them comparable; 2) the aggregation of the standardized indicators into a single composite index.

5.   What is meant by the "polarity" of an indicator?

It is the "direction" of the relationship between the indicator and the phenomenon. A positive polarity means that as the indicator increases, the phenomenon improves (e.g., well-being); a negative polarity means that as the indicator increases, the phenomenon worsens (e.g., crash rates).

6.   Why was a negative polarity set for the crash indicators?

Because it is more intuitive to associate a negative value with a disadvantageous situation or higher risk for a province, making the rankings easier to read and interpret.

7.   What is a "formative" measurement model?

It is a model where the indicators are seen as "causes" of the phenomenon. A change in a single indicator does not necessarily imply a change in all the others. This model was assumed for the synthesis methods used.

8.   What is the MPI (Mazziotta-Pareto Index) method?

It is an alternative method for creating composite indices "non-compensatory," meaning it penalizes the imbalance between the values of the indicators (a deficit in one indicator cannot be easily offset by a surplus in another).

9.   Why were some synthesis methods, like the "Rankings" method, excluded?

Because they were not considered suitable for the phenomenon of road crashes. The Rankings method, for example, does not account for outliers, which can be significant in this context.

10.   What is the ultimate advantage of using a composite index?

It simplifies analysis and communication by providing a one-dimensional measurement of a complex phenomenon and allowing for an easy ranking of provinces based on their overall risk level.

Edited by: Marco Broccoli (broccoli@istat.it) e Silvia Bruzzone (bruzzone@istat.it)