19 MAY 2021

USE OF OPEN STREET MAP FOR ACCIDENT INVESTIGATION ON THE ROAD AND MOTORWAYS NETWORKS.
UPDATING YEAR 2017.

1. Road Safety performance Indicators and Big data use

Road Safety Performance Indicators (RSPI) give a multidimensional approach for accident investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and Big Data (BD) sources the result gives new elements on planning infrastructure solution, applying policies to reduce deaths and serious injuries, reducing social costs on collectivity and estimating efficiency and effectiveness of safety initiatives.

Preventing road trauma on public roads is a core responsibility for government, its agencies and stakeholder. It requires a common and shared responsibility. The scale of the road safety challenge and the diversity of the effects of road traffic injury underline the importance of exploring synergies among the decision makers of the road network.

Nowadays there is a clear information bias as regards the appropriate reference denominators to be placed as basis in construction of statistical indicators linked to road accidents. Resident population is used as a common proxy for exposed at risk in a specific geographical area, but not always an appropriate solution, especially in the light of the seasonal nature of road accidents and concentration, in some periods of the year, in specific locations. Vehicle fleet (Automobile Club of Italy - ACI) can be another administrative source that gives a more accurate information, but the characteristic of the phenomenon implies a deductible distortion on measures due to the mobility of road users.

The length of the road network (from Open Street Map) gives for sure a consistent first set of information concerning the different territories.

This information is not available from official statistics at national level, although there are archives and detailed road graphs for each municipality, province and region, a harmonized and systematic national road registry has not been established.

The first output of the project, in line with the process of modernization of Istat statistical production, is the focus on the exploitation of existing administrative sources, the scouting of new sources and the analysis of integrated and auxiliary data. The basis of the renewal in the statistical production is to upload any source integration; even any new technique implemented and applied methodology. Every small change that overall effect becomes a process of improvement of the quality of the statistical information provided by Istat.

The aim of project is to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map. Although the product represents a first result, the final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, taking into account the different exposure to risk.

2. The use of GIS systems for a graphical representation

GIS is a geographic system designed to capture, store, manipulate, analyse, manage, and present spatial or geographic data. GIS applications are tools that allow users to analyse spatial information, edit data in maps, and present the results of all these operations. In order to relate information from different sources, GIS uses spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS has begun an alternative frontier on producing statistical information. Any variable that can be located spatially using an x, y, and z coordinates representing, longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

Join attributes by location is the algorithm that takes an input vector layer and creates a new vector layer that is an extended version of the input one, with additional attributes in its attribute table. The additional attributes and their values are taken from a second vector layer. A spatial criteria is applied to select the values from the second layer that are added to each feature from the first layer in the resulting one (Chart 1).

Chart 1 – Graphical representation of the join attributes by location algorithm

2.1 Istat Census Map localities

The Istat Census Map localities used with the GIS system are provided for the following administrative units:

- Regions
- Provinces
- Municipalities (yearly updated)
- Localities (only 2011 Census)

An upgrade of Census Map localities shapes referred to 2011 to municipalities 2016 has been done. It has been built a link table with the aggregation of all 8090 local administrative units territory, at 2011, to the 7998 municipalities at 2016, included in the Italian territory.

The choice of the localities shapes is due to the harmonization need of the complete roads graph to the "road type classification" used by the road accidents survey.

The localities classification includes 4 voices:

1. Urban areas
2. Small inhabited areas
3. Productive areas
4. Wide spread houses

As regards the definitions, the Istat[1] descriptions are reported below.

Localities - A more or less wide area, usually known by name, on which one or more houses are grouped or scattered; there are two types of localities: inhabited localities and productive locations. The borders of the inhabited localities (center and inhabited area) are the external limit of the buildings placed at the edges of a grouping of at least fifteen buildings. The confines of the localities include gardens and other accessories areas of the considered buildings, non-built neighboring areas not included (such as fields with or without crops). Buildings located at a distance of more than 70 meters, within built-up areas and 40 meters for small inhabited areas excluded. If the buildings included in the new locality are adjacent or in proximity (to a max distance of 140 meters for urban areas and 60 meters for small inhabited areas) of transport infrastructures or hydrographic limits, the location border will be extended to the middle of these elements.

Urban areas - Aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities (school, public office, pharmacy, shop or similar), detecting a social life and, generally, also a place of collection for the inhabitants of

---

[1] Istat - Basi territoriali e variabili censuarie https://www.istat.it/it/archivio/104317
Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011

https://www.istat.it/it/files//2013/11/Descrizione-dati-Pubblicazione-2016.03.09.pdf the neighboring areas. The places of tourists meetings, houses, hotels and similar used for the vacation, inhabited seasonally, are considered as temporary inhabited centers too.

Small inhabited areas is an area without the place of collection, characteristic of the urban area. It is based on a group of at least fifteen contiguous and near buildings, with at least fifteen families, with roads, paths, squares, farmyards, small gardens and similar, as long as the distance between the buildings does not exceed thirty meters and it is lower than the distance between the center and the nearest of the houses clearly scattered.

Productive areas - Extra-urban area not included in the centers or residential areas with more than 10 local units, or with a total number of employees' upper than 200. The local units are contiguous or close, with roads, squares or similar, or anyway in a continuous line, not exceeding 200 meters; the minimum area must be 5 hectares.

Wide spread houses - Houses scattered in the municipal territory at a distance not enough to constitute a built-up area.

2.2 Open Street Map and road arch type

Open Street Map (OSM) [2] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps are created using the data recorded by portable GPS devices, aerial photographs and other free sources. Most of the Android and iOS GPS navigation software on portable devices are powered by OSM as WisePilot, Maps.me, NavFree, Scout etc.

The Open Street Map vector layers, used in this work, daily updated and free downloadable data, are the following:

- Road graph;
- Point of traffic (POT);

Added shape:

- Buildings;
- Use of the land;
- Natural;
- Places;
- POWF (Point of Worship);
- POIS (Point of interest);

[2] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: http://www.openstreetmap.org/about

- Railways;
- Transport;
- Water;
- Water ways.

Although OSM it is an Open Source tool based on information from a community, the product provides data to be considered reliable and consistent, so much that the major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, etc.

Table 1 contains the list of different type of road arch by Open Street Map.

Table 1 - OpenStreetMap road arch classification (a)

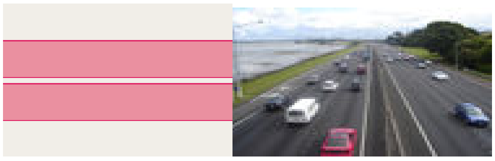
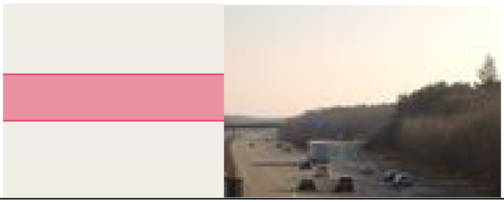
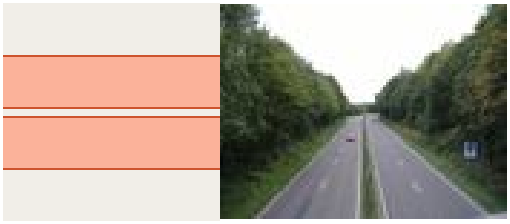
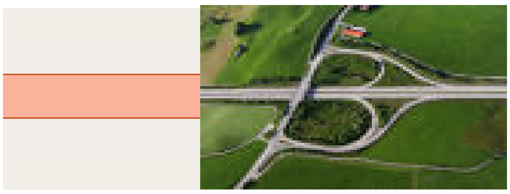
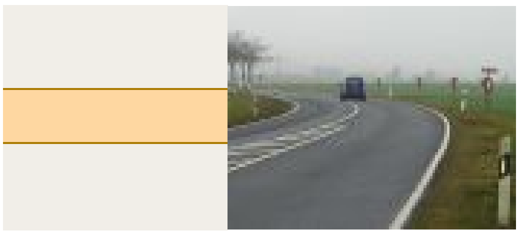
	**Motorway** Free or by toll payment highways, including motorway connections such as the Milan ring road and the Grande Raccordo Anulare of Rome. Equivalent to Freeways, Autobahns, etc ...
	**Motorway Link** The link roads (sliproads/ramps) leading to/from a motorway from/to a motorway or lower class highway. Normally with the same motorway restrictions.
	**Trunk** Roads type between Motorway, Motorway connections and Primary road. The junction section of a motorway-ring road that leads to the city center can also be classified as trunk. Some special primary road called in Italian “Superstrade” with two lanes, could be included in this category. In “trunk” are included extra-urban roads with only one lane in each direction, without crossings, with exits and acceleration and deceleration lanes. They do not include intersections and roundabouts.
	**Trunk Link** The link roads (sliproads/ramps) leading to/from a trunk road from/to a trunk road or lower class highway.
	**Primary** Roads of national and regional importance not classified as motorways, trunks, or their link. They connect the main cities to each other. Usually, they are classified as SS (Main Roads) or SR (Regional Roads), however there are some exceptions, such as in small mountain towns where the main road crosses the village but the primary road is a ring road of modern construction, around the inhabited center. In urban areas, they represent the external ring of the city (e.g. Milan) and are normally classified as Viali.
	Primary link The link roads (slip roads/ramps) leading to/from a primary road from/to a primary road or lower class highway.

Table 1 - OpenStreetMap road arch classification (continued) (a)

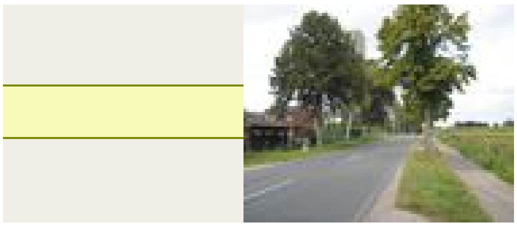
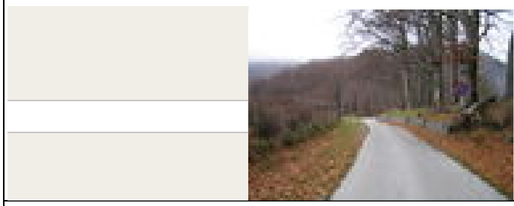
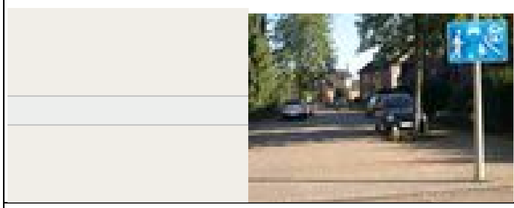
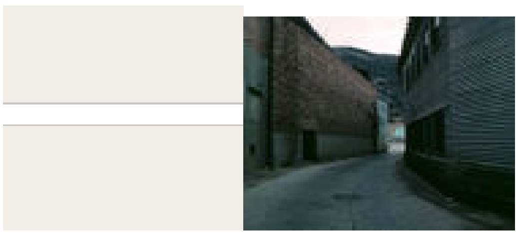
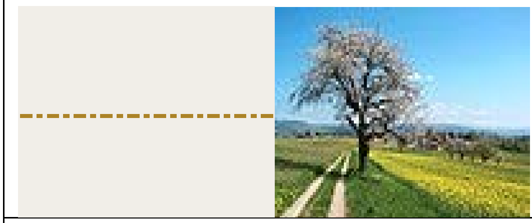
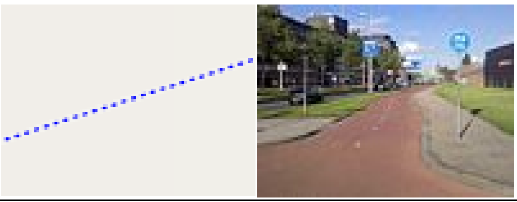
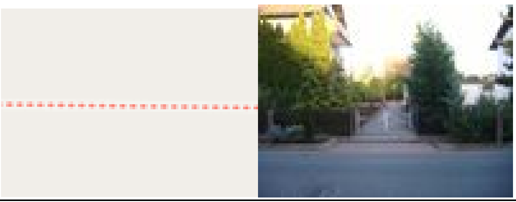
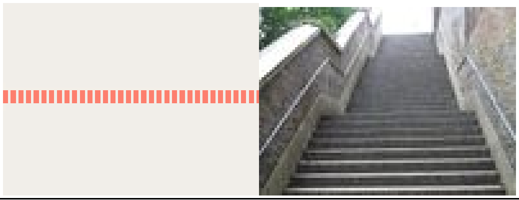
	**Secondary** Roads of regional and provincial importance. They link together the main municipalities of a region. They are, usually, classified as SP (provincial roads) but there are some exceptions. In urban areas, they are normally classified as important streets with two lanes in each direction.
	Secondary link The link roads (sliproads/ramps) leading to/from a secondary road from/to a secondary road or lower class highway.
	**Tertiary** Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
	Tertiary link The link roads (sliproads/ramps) leading to/from a tertiary road from/to a tertiary road or lower class highway.
	**Unclassified** Classification for some extra-urban road. In urban areas, they are used to reach close destinations and the traffic flow is lower than in higher classification roads. They often connect villages and hamlets.
	**Residential** Roads in a residential area, which serve as an access to housing, without function of connecting settlements. Often lined with housing.
	**Living Street** Residential road where pedestrians have legal priority over cars, speeds are kept very low and where children are allowed to play on the street (category not frequently present in Italy, anyway, if used, the signal of residential area is installed).
	**Pedestrian** Pedestrian areas (roads or squares in urban areas), accessible mainly or exclusively to pedestrians.
	**Service** Access roads or internal service areas, beaches, camping, industrial areas, shopping centers, residences, parking places, landfills, installations, etc.

Table 1 - OpenStreetMap road arch classification (continued) (a)

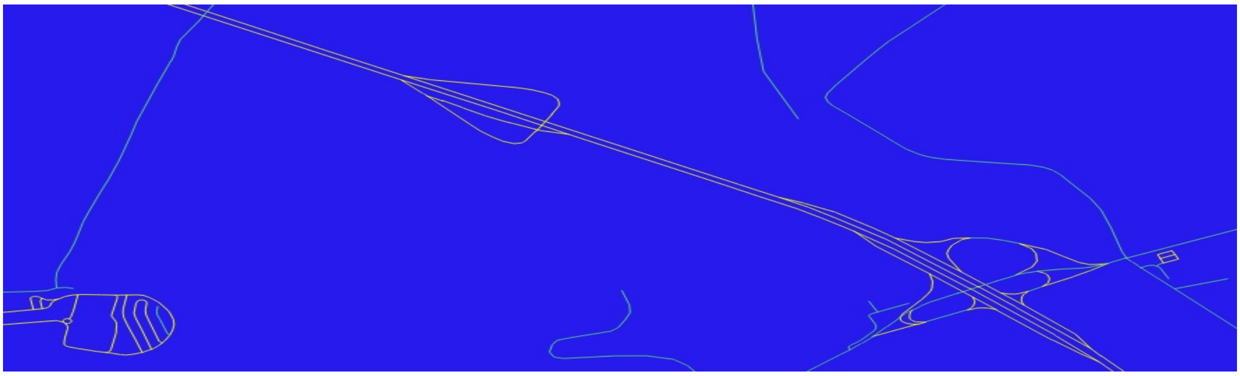
	**Track** Roads for mostly agricultural or forestry uses. To describe the quality of a track, are often rough with unpaved surfaces.
	**Bridleway** Paths for horse riders (usually they are pedestrian area with the access allowed to horses too).
	**Cycleway** Cycle paths on dedicated carriageway, mainly or exclusively for cycling tourism.
	**Footway** Paths mainly/exclusively for pedestrians. This includes walking urban tracks, paths in a public park and footpaths also if managed and not maintained at natural status.
	**Path** Paths not structured for a public use
	**Steps** Stairs in steps, exclusively accessible by pedestrians
	**Unknown** Not classified

(a) Road Archs by Open Street Map – update 16/2/2017.

An innovative method of measuring the length in meters of the road graph is given by the information on the number of carriageways of each road arch of OSM.

The snapshot in Chart 2 shows how the yellow arches represent the one way archs, while the green ones the two-ways. In the future, in order to provide even more detailed information, the use of the number of lanes containing each carriageway is being evaluated.

Chart 2: The OpenStreetMap information on carriageway numbers



2.3 Classification of OSM road arches and road accidents localization

The arch road types selected to calculate the indicators analysed are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track_grade, bridleway, cycle way, footway, path, steps are not object of the survey definition.

To build road accidents indicators, with denominator represented by the arches length in Open Street Map, since the last edition, we built a “bridge matrix” between road categories, classified by functional road type, used by OSM, and the categories linked to the roads holder, used by Istat road accidents survey. This methodology can be defined as **a systematic classification technique**.

The systematic classification of road arches (Open Street Map), classified in the categories used by Istat, has been modified, in respect of the first release of the experimental statistics. A new **analytical classification** has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total (OSM), to the Istat classification groups See table 2). The operational criterion applied provides the roads classification, through the textual analysis of the Name and Reference attributes, according to the different classes of road arch and spatial attribution of the location type.

In order to use 2017 data, improved by the new technique and to guarantee the comparability to 2016 data, already disseminated, the new 2017 indicators were also completed by a recalculation of the provincial rankings, adopting the 2016 systematic classification (Chart 3-8).

Chart 3: Motorways. Year 2017 (Istat: analytical classification).

Chart 4: Main Roads. Year 2017 (Istat: analytical classification)

Chart 5 Regional Rural Roads. Year 2017 (Istat: analytical classification)

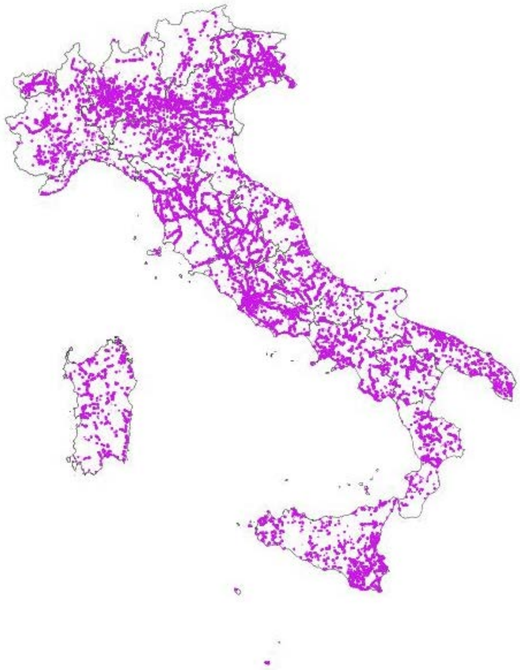


Chart 6: Rural Roads. Year 2017 (Istat: analytical classification)

Chart 7: Local and other rural roads Year 2017 (Istat: analytical classification)

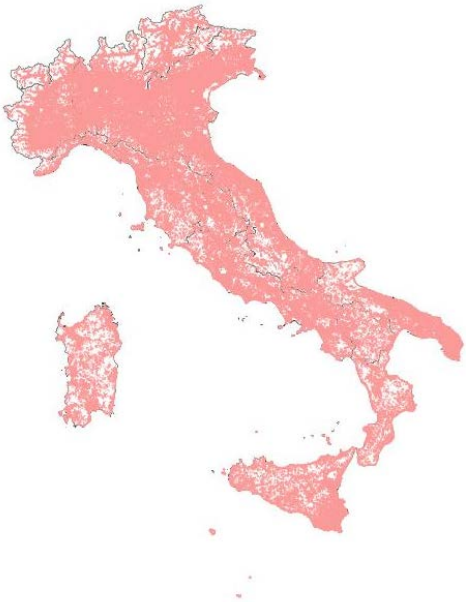


Chart 8: Urban Roads. Year 2017 (Istat: analytical classification).

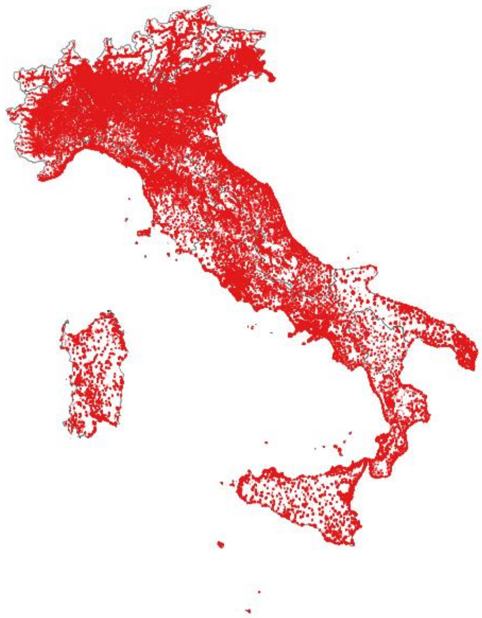


Table 2 – Bridge table between roads arches classification by OSM, localities and road type (a)

Road Archs classification by OpenStreetMap	Localities at Census 2011			
	Urban areas + Small inhabited areas		Productive areas + Wide spread houses	
	Road Localisation by Road accidents survey			
	Motorways	Urban Roads	Motorway	Rural Roads
Motorway	X		X	
Trunk	X		X	
Primary		X		X
Secondary		X		X
Tertiary		X		X
Unclassified		X		X
Residential		X		X
Living Street		X		X
Motorway Link	X		X	
Trunk Link	X		X	
Primary Link		X		X
Secondary Link		X		X
Tettiary Link		X		X
Service		X		X
Unknown		X		X

(a) Istat computing.

3. The indicators on road accidents

3.1 Road accident indicators in relation to the extended road, resident population, vehicle fleet

The proposal to calculate new road accidents indicators was born to provide, new and more suitable measures, for the risk and probability of being involved in a road accident. Even if Resident Population and Vehicles fleet are both denominators considered, at present, *proxy* for the risk exposure to the road accidents, it is clear that they present some critical points.

As regards the data updating (numerators of indicators) used to build indicators on the road accidents risk, they refers to Istat "road accidents survey", concerning all road accidents resulting in deaths (within the 30th day) or injuries in 2017, involving at least a vehicle circulating on the national road network and documented by a Police authority[3].

As regards denominators used, they are the Resident Population, Vehicle Fleet and the length in meters per carriageway.

The Resident Population[4], not always is an appropriate solution to build road accident indicators, because the seasonality of accidents and the concentration in some specific places during the year. The Vehicles fleet by registration province (Automobile Club of Italy – Public Register of Vehicles PRA)[5] provides more suitable information than the population, but it does not consider the mobility of users through the country.

The length in meters of carriageway by road direction of road arch (by OSM) [6] provides, instead, a set of more coherent information referred to different areas, because independent by the seasonality and the mobility users influence.

This information is not available at national level, since, although archives and detailed road graphs by single municipality, province or region exist, a national and harmonised road register has been not created yet.

3.2 Road accident indicators: weighting with Traffic Points (PoT) from the Open Street Map

The length in meters of carriageway per direction of the road arch (from Open Street Map), used to calculate the indicators for experimental statistics, provides a first coherent set of information relating to the various territories. Although the product constitutes a first usable result, the project's most ambitious goal is to estimate the real traffic flows (vehicles / Km) on the national

[3] Survey on road accidents resulting in death or injuries https://www.istat.it/it/archivio/4609

[4] Resident Population - 31/12/2017.

[5] ACI Vehicle fleet (Automobile Club of Italy) all motorized vehicles except the trailers on 31st December 2016.

[6] GIS computing (Census Map + Open Street Map road graph at 1st January 2017) expressed by length in meters per carriageway.

road network. This would make it possible to calculate the probability of being involved in a road accident and therefore of real exposure to the risk of accident.
To fill this information gap, other additional information has been obtained from the rich source Open Street Map, in particular, data on the traffic points detected on the road arches (PoT Point of Traffic). In fact, a monthly information, downloadable from OSM, is available and the data refers to points over an arch, in which is detected an intensity of traffic. The proposed new road accident indicators, "weighted" with the information on traffic intensity, was built considering, as a discriminating element, the kilometers of carriageway with the presence of a traffic point on the arch.
The authors calculated the length of roads, considering the extension, in meters, of carriageway with arches on which including the presence of traffic points. The relative frequency (f) of the road length with the presence of traffic points on an arch (by province and type of road) and the complementary frequency (1-f) of arches without point of traffic has been calculated too.
The number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, excluding the effect on accidents caused by the presence of traffic points on the roads has been processed. The calculation was calculated by multiplying the value of the indicators for road length and the frequency (1-f) referred to arches without traffic points on the roads of the province.

With reference to the notation used for the calculation, the individual items are defined as follows:

LA Tot $_{i,j,k}$ = Total Length of Arches in meters in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA PoT $_{i,j,k}$ = Length of Arches in meters with presence of Points of Traffic (PoT) in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA Tot $_{i,k,l}$ = Total Length of Arches in meters in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

LA PoT $_{i,k,l}$ = Total Length of Arches in meters in the province with presence Points of Traffic (PoT) in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

The total length of the national roads, referred to arches length, with the presence of Points of Traffic (PoT) or in total, is given by the following expressions:

Total Length of Arches with Points of Traffic (PoT):

$$\text{LA PoT}_{tot} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA PoT}_{i,j,k} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA PoT}_{i,k,l}$$

$$\text{LA Total}_{tot} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA Totale}_{i,j,k} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA Total}_{i,k,l}$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by functional class of Open Street Map Arch (j) and Locality type (k) out of the total length is given by the following expression:

$$\text{p LA PoT}_{i,j,k} = \frac{\text{LA PoT}_{i,j,k}}{\text{LA Tot}_{i,j,k}} * 100$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by Locality type (k) and Type of road by Istat classification (l) out of the total length is given by the following expression:

$$\text{p LA PoT}_{i,k,l} = \frac{\text{LA PoT}_{i,k,l}}{\text{LA Tot}_{i,k,l}} * 100$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), at national level, out of of the total length, is given, at last, by the following expression:

$$\text{p LA PoT}_{tot} = \frac{\text{LA PoT}_{tot}}{\text{LA Total}_{tot}} * 100$$

where **i**= 1,…, 103, 108, 109, 110, 111 (Istat Code of province)

j= 1,.....,27 (Functional class of the Open Street Map Arch)

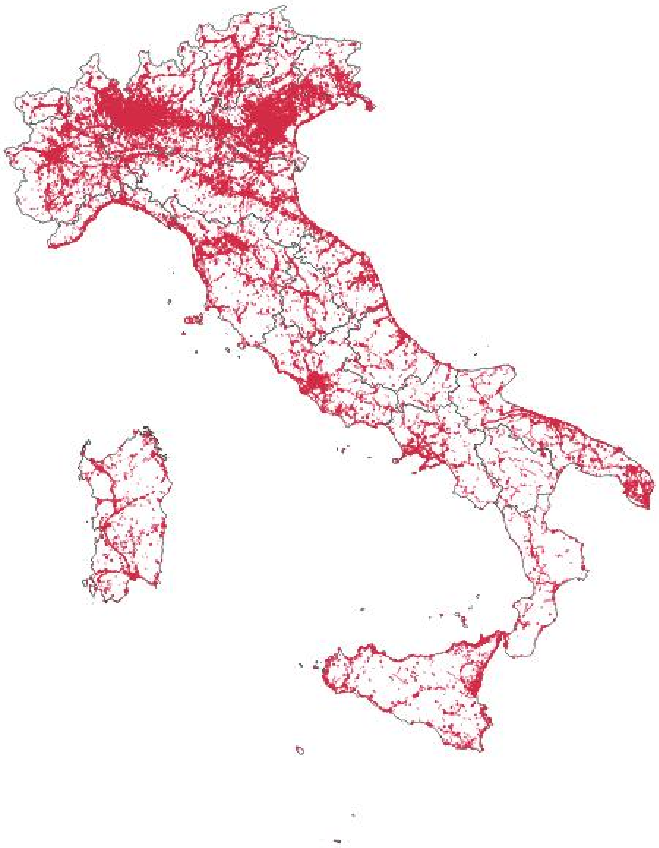
1=Motorway; 2=Motorway_Link; 3=Trunk; 4=Trunk_Link; 5=Primary; 6=Primary_Link; 7=Secondary; 8=Secondary_Link; 9=Tertiary; 10=Tertiary_Link; 11=Unclassified; 12=Unknown; 13=Residential; 14=Living_Street; 15= Pedestrian; 16=Service; 17=Path; 18=Steps; 19=Track; 20=Track_Grade1; 21=Track_Grade2; 22=Track_Grade3; 23=Track_Grade4; 24=Track_Grade5; 25=Bridleway; 26=Cycleway; 27=Footway

k=1,.......,4 (Type of locality: 1=Urban areas, 2=Small inhabited areas, 3=Productive areas, 4=Wide spread houses)

l= 0,……,9 (Type of road by Istat classification: Regional roads outside urban area (0), Urban roads (1), Rural roads outside urban area (2), Main roads inside urban area (3), Local rural roads (4), Rural roads outside urban area (5), Main roads outside urban area (6), Other roads outside urban area (8), Regional roads outside urban area (9))

In 2017, there are about 220,000 Traffic Points on the road arches of the national road network and almost 68,000 km of carriageway, considering the extension of all the arches that contained at least one traffic point. This extension represents, in Italy, 6.3% out of the total network. The types of roads, according to the Istat classification, which have the highest percentages of traffic points are motorways (31.7%), national roads in the built-up area (30.7%) and regional roads in the built-up area (30.2 %).
With this criterion, in fact, with the same number of accidents and road kilometers extension, for a similar category of road in two different provinces, for example, the presence of PoT modifies the risk of road accidents, with disadvantage of road sections with less traffic flow, thus they result with an higher danger in terms of vehicles / km (Chart 9).

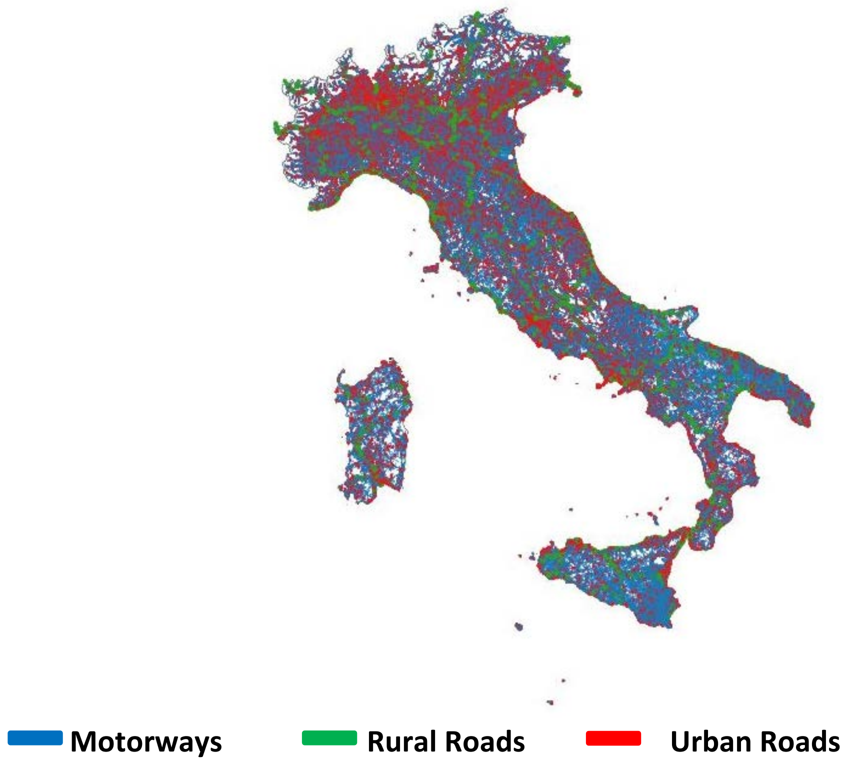
Chart 9: OpenStreetMap Road Arches with Point Of Traffic. Year 2017



3.3 Road accidents indicators computing: main results

The indicators have been updated with 2017 data, comparing the values to the version already published referred to 2016 data and, with reference to the rates on the road length, they were also calculated, for 2017, with the analytical roads reclassification, by road type (Chart 10).

Chart 10: Motorway, Rural and Urban Roads. Year 2017 (Istat: analytical classification).



The proposed indicators refer, in fact, to the ratio between road accidents, deaths, injuries and vehicles to the road length, considering the additional information on points of traffic, the resident population and the vehicle fleet.

The road accidents indicators out of the road length provide a measure of the number of accidents, vehicles involved, dead and injured per 100 kilometers of carriageway in the province.

Table 2.14 (See Data Tables) contains the values of the "Rates for road accidents, vehicles involved, killed and injured per 100 kilometers of carriageway by Italian province (classification 2016) - Year 2017", Table 2.15 refers, on the other hand, to the same 2017 indicators according to the new 2017 road arches classification. As regards the results contained in the Table 2.15, a maximum exposure to the risk of accidents and number of vehicles involved, for motorways and urban roads, is stated mainly in main cities. Milan records a maximum for accidents and vehicles rates per 100 km of carriageway, on Motorways (respectively 144.69 and 313.54 for the 2017 road arches classification) and Genoa on Urban roads, for accidents and vehicles per 100 km

(152.91 and 256.66). On Rural roads, on the contrary, the medium-sized provinces result most affected: Monza and Brianza has the maximum value with 62.98 accidents per 100 km and 125.67 vehicles.

As regards the road accidents deaths on Motorways, Urban and Rural roads for 100 km of carriageway (2017 arches classification) the highest values are recorded respectively for: Modena (5.72), Ravenna (1.19), Naples (1.68). The indicators referred to injured on Motorways, Urban and Rural roads for 100 km of carriageway (classification arches 2017) show high values for Milan, Genoa and Monza and Brianza (222.04; 189.04; 98.48).

The set of road accident indicators, calculated out of the vehicle fleet amounts, measures the number of accidents, vehicles involved, deaths and injuries per 100,000 vehicles registered in the province.

In terms of road accidents and injuries by type of road, compared to the vehicle fleet, the indicators reach the maxim value on Motorways in Savona (58.82 and 104.53 per 100,000 vehicles in vehicle fleet), in Genoa for Urban roads (697, 96 and 862.84) and in Grosseto for Rural Roads (157.63 and 238.62).

The indicators, calculated with the ratio between vehicles involved in road accidents and vehicle fleet (per 100,000), show a maximum level on Motorways in Bologna (114.27), Genoa for Urban area (1171.51) and Pisa for rural area (268.82).

Road mortality (per 100,000 vehicles registered) is high in Verona on the Motorways (2.76), in Pistoia on Urban roads (7.52) and in Foggia on Rural roads (14.24).

Concerning key results, on the motorway sector, the low incidence of vehicles registered in the province with the presence of important infrastructural nodes and seasonal factors amplifies the distortion of the indicators distribution. In urban areas, commuting is not highlighted in the construction of the vehicle fleet. In the rural area, this category of indicators does not show the presence of a dense network of consular roads on the territory (See data tables - Table 2.16).

The **set of road accident indicators, calculated using the resident population**, measures the number of accidents, vehicles involved, deaths and injuries per million residents in the province.

Concerning the road accidents rates, vehicles and injured by road type, per resident population, they present high values on Savona Motorways (respectively 565.13; 1087.07 and 1004.28 per 100,000 inhabitants for accidents, vehicles and injured). On Genoa Urban roads (respectively 5676.03, 9527.11 and 7016.93 per 100,000 inhabitants for accidents, vehicles and injured) and on Grosseto Rural roads (respectively 1462.81; 2480, 03 and 2214.47 per 100,000 inhabitants for accidents, vehicles and injured).

Road death rates (per 100,000 inhabitants) are highest in Verona on Motorways (23.84), in Ravenna on Urban roads (63.88) and in Foggia on Rural roads (95.95).

Some notes have to be specified. The resident population use could introduce an interpretation bias, in fact, it does not include always the real number of users exposed to risk of accidents (this is the case of the motorway network nodes). Besides, the presence of port areas, transit areas and productive settlements in urban areas can show an exposure to risk of a much higher number of individuals than those included among residents. Even for rural roads the real extent of the phenomenon is not taken into account. The indicators in these cases, therefore, could be biased or amplified (See data tables Table 2.17).

The **Road accidents indicators by road length "weighted" by the PoT** (Point of Traffic) information measures the number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, "adjusted " by the effect on accidents of the presence of traffic points on the roads. The ratios are based on the product between the value of the indicators by road length (See Table 2.15) and the complementary percentage (1-p) to the presence of arches with traffic points on the roads of the province (See Table 2.15 T).

The table shows as often the same province presents different positions in the ranking of the indicator by road length, compared to the ranking of indicators adjusted by traffic effect. This is the case of the road accidents on the highways in Genoa and Milan, e.g, the position of Genoa, in fact, reveals a disadvantaged position, in respect of Milan, considering the real danger of the roads of the province with traffic.

The rates for road accidents and injured per 100 kilometers of carriageway, weighted with information Point of Traffic, present a maximum value on the Motorways in Genoa (85.58 and 134.92 respectively for accidents and injured) and Milan for vehicles involved (177.89). In Genoa, on Urban roads for road accidents and injured per 100 kilometers of carriageway (values respectively 116.88 and 144.49 for accidents and injured), in Monza and Brianza on Rural roads, for accidents, vehicles and injured (values equal to 52.79, 105.34 and 98.48 for accidents, vehicles and injured).

As regards the mortality rates, Verona presents the highest value on Motorways (4.35 deaths per 100 kilometers of carriageway, without traffic), Ravenna on Urban roads (1.02) and Naples on Rural roads (1.60) (See data tables Table 2.15 T).

Finally, the set of 2017 indicators is completed by a group of composed and synthetic road accidents indicators. The method used is the arithmetic mean (z-scores) MZ method (Software Ranker - See Paragraph 4 and Table 2.18).

The values of indices have been calculated according to the MZ method for road arches; vehicle fleet and population also using the old systematic classification of arches, such as the release used last year, in order to make correctly a comparison between the years 2016 and 2017 (See data tables Table 2.19).

It should be noted that the Ranker software, used for the calculation of synthesis indicators, requires to set the “polarity” of the indicator in order to define a values ranking, attributing a specific meaning with a + or - sign.

The 2016 synthesis indicators were calculated by setting the polarity with a positive sign, to highlight the greater danger associated with road safety in the provinces, if a high value of the indicator stated.
For 2017 data, an additional element was included, the "points of traffic factor". The new weights used for the calculation of the indicators for road length are represented by the complementary percentage (1-p) of the proportion of road extensions with traffic points (p). This method is applied with the aim of excluding the traffic effect and highlighting the consequences of drivers' driving behavior and damages in infrastructures. It is considered more intuitive, in fact, to consider the negative polarity of the indicator, linked to dangerous situation in the provinces.
In order to make a direct comparison between the synthetic indicators referred to the years 2016 and 2017, therefore, the polarity was reversed for the 2016 indices already disseminated (See data tables Table 2.19).

4. The computing of synthesis indicators

To realize a ranking and a classification of indicators, a generalized tool, developed by Istat, was used. Two generalized tools are available for the analysis and benchmarking of results produced by different composite indicators: RankerTool is a desktop and i.Ranker web application (Appendix A).

Both, with few differences, allow the user to:

- acquire, in standard format (csv or .xls), the values of the indicators available for each entity, already calculated and normalized;
- calculate, for each entity, one or more methods among those implemented;
- display the values and rank results for each method, both in tables and graphic way;
- compare the rankings using the different methods.

The Ranker tool used give the possibility to choose different methods.

The methods proposed are: the Average method of standardized values (MZ), the MR method - relative indices (IR), the MPI method - Mazziotta-Pareto Index (De Muro et al. 2010), the Graduations Method (MG) and the Wroclaw Taxonomic Method (MTW). The last two options were excluded a priori, as they were not considered suitable for ranking indicators on road accidents and representing the phenomenon under study. The ranking method, in fact, does not take into account the outliers, while the Wroclaw taxonomic method is based on the distance from an ideal unit in the Euclidean space.

The steps on the computing process is divided in two phases:

- **standardisation** of elementary indicators, the standardization aims to make the indicators comparable as they are often expressed in different units of measurement and may have different polarities. Therefore, it is necessary to bring the indicators to the same standard, reversing the polarity, where necessary, and turning them into pure, dimensionless numbers;

- **aggregation of standardized indicators**: the combination of all the components to form the synthetic index (mathematical function), after the standardisation by Ranker tool[7].

For the selection of the method to be used for synthetic indexes computing, a **robustness test** was carried out and an **influence analysis** was done by applying the COMIC[8] software (COMposite Indices Creator), through synthesis methods and the evaluation of their consistency.

Through the COMIC software, it was possible to make a comparison between the degrees of robustness for the main applicable methods; in particular, table 6 shows the data on the mean and standard deviations of the *shifts* for different methods.

The results of the influence analysis and robustness tests data show that the best methods to be used, among the different alternatives, are the MZ method - arithmetic mean (z-scores) and MPI - Mazziotta -Pareto Index.

The **MZ method - arithmetic mean (z-scores)**, well known method that allow an easy interpretation of results for all common users, has been selected for the presentation of the data contained in this research.

Table 7 – Summary of means and standard deviation of the *"shifts"* (a)

Method for Synthex Indicators	Mean of the mean of the shift	Standard deviation of the mean of the shift	Mean of the Standard deviation of the shift	Standard deviation of the Standard deviation of the shift
Index Mean 0-1	2,092	0,828	2,152	0,857
MZ aritmetic mean z-scores	2,224	0,604	2,356	0,631
Jevons static	2,671	1,416	2,752	1,221
MPI	2,183	0,667	2,440	0,938
MPI correct	2,136	0,862	2,288	1,090
IMG	2,006	1,098	2,530	2,057

(a) Results of the robustness test, using the COMIC software - Synthesis indicators for road accidents rates by road arch, vehicles fleet and resident population. Year 2016

The application of different weighting criteria leads to very divergent results. The analysis according to the road infrastructures allows purifying a component of mobility of the phenomenon. The seasonal factor due to a more objective measurement also improves the concept of exposure to the risk of being involved in a traffic accident.

Table 8 shows the covariance values, referred to the variation of each variable contained in the matrix in respect of all others.

---

[7] The methodological note and the use guide are available in Appendix A of this document and in the links:
http://www.istat.it/en/files/2014/03/RANKER-manuale.pdf
https://i.ranker.istat.it/wr_guida.htm
https://i.ranker.istat.it/wr_guida_notametodologica.htm

[8] COMIC (COMposite Indices Creator) https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it/analisi/strumenti-di-analisi/comic

Table 8– Covariance Matrix between computing indicators: results obtained by Ranker application. 2017 data

Ranks	Road Arch	Vehicles Fleet	Resident Population
Road Arch	1,0000	0,5588	0,4550
Vehicles Fleet	0,5588	1,0000	0,9331
Resident Population	0,4550	0,9331	1,0000

The values included in table 8 show, in fact, that the risk to be involved in a road accident, within the province of residence (0.4550) or within the vehicle registration province (0.5588) is significantly lower if compared to the indicator out of the length road arch, where the accident occurred.

The road accidents indicators, referred to the road length by province, therefore, allows to obtain a result closer to the road accident risk measures, using traffic flows data. The last data would represent, in fact, the best and correct denominator for road accidents rates.

Table 9 shows the values of the covariance, referred to the variations of the different classification criteria of the road arches.

Table 9 - Covariance matrix for the classification criteria of road arches on 2017 data

Ranks	Systematic classification	Analytical classification	Classification with PoT
Systematic classification	1,0000	0,9940	0,9793
Analytical classification	0,9940	1,0000	0,9828
Classification with PoT	0,9793	0,9828	1,0000

The table shows as the analytical classification criteria for road arches leads to a qualitative improvement in the attribution of the events, according to the Istat road accidents survey variable (0.9940), without significantly altering the two provincial distributions. The inclusion of the traffic factor allows reading the road accidents components net of other factors, such as the infrastructural and behavioral. It leads, moreover, to a more objective evaluation criterion of the phenomenon (0.9828). This measurement makes it more evident the identification of actions to improve road safety by the stakeholders and in the prevention actions by the law enforcement agencies in charge of road control.

APPENDIX A

Ranker tool for a composite index computing

A1. Computation and evaluation of composite indices [9]

A *composite index* is a mathematical combination (or aggregation as it is termed) of a set of indicators [1] that represent the different dimensions of a phenomenon to be measured.

Constructing a composite index is a complex task. Its phases involve several alternatives and possibilities that affect the quality and reliability of the results. The main problems, in this approach, concern the choice of theoretical framework, the availability of the data, the selection of the more representative indicators and their treatment in order to compare and aggregate them.

In particular, we can summarize the procedure in the following main steps:

1. *Defining the phenomenon to be measured*. The definition of the concept should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators. Also the *model of measurement* must be defined, in order to specify the relationship between the phenomenon to be measured (concept) and its measures (individual indicators). If causality is from the concept to the indicators we have a *reflective model* – indicators are interchangeable and correlations between indicators are explained by the model; if causality is from the indicators to the concept we have a *formative model* – indicators are not interchangeable and correlations between indicators are not explained by the model.
2. *Selecting a group of individual indicators*. The selection is generally based on theory, empirical analysis, pragmatism or intuitive appeal. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility and so on. The selection step is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. However, the selection process also depends on the measurement model used: in a reflective model, all the individual indicators must be inter-correlated; whereas in a formative model they can show negative or zero correlations.
3. *Normalizing the individual indicators*. This step aims to make the indicators comparable. Normalization is required before any data aggregation as the indicators in a data set often have different measurement units. Therefore, it is necessary to bring the indicators to the same standard, by transforming them into pure, dimensionless, numbers. Another motivation for the normalization is the fact that some indicators may be positively correlated with the phenomenon to be measured (positive *polarity*) [2], whereas others may be negatively

[9] RankerTool desktop software (http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ranker)
i.Ranker web application (https://i.ranker.istat.it)

correlated with it (negative *polarity*). We want to normalize the indicators so that an increase in the normalized indicators corresponds to increase in the composite index. There are various methods of normalization, such as *re-scaling* (or Min-Max), *standardization* (or *z*-scores) and 'distance' from a reference (or *index numbers*).

4. *Aggregating the normalized indicators*. It is the combination of all the components to form one or more composite indices (mathematical functions). This step requires the definition of the importance of each individual indicator (weighting system) and the identification of the technique (*compensatory* or *non-compensatory*) for summarizing the individual indicator values into a single number [3]. Different aggregation methods can be used, such as *additive methods* (compensatory approach) or *multiplicative methods* and *unbalance-adjusted functions* (non-compensatory or partially compensatory approach) [4].
5. *Validating the composite index*. Validation step aims to assess the robustness of the composite index, in terms of capacity to produce correct and stable measure, and its discriminant capacity (*Influence Analysis* and *Robustness Analysis*).

A2. The Synthesis Methods

The synthesis methods available on i.ranker are based on the assumption of a formative model:

- Mean of standardised values (MZ);
- Mean of relative index (MR);
- Mazziotta-Pareto index (MPI^+/MPI^-);

Mean of the standardised values (MZ)

Given the matrix **X**={x_{ij}} of *n* rows (units) and *m* columns (indicators), transformation matrix **Z**={z_{ij}} is built, with:

$$z_{ij} = \begin{cases} \dfrac{(x_{ij} - M_{x_j})}{S_{x_j}} & \text{if the indicator j has positive polarity} \\ -\dfrac{(x_{ij} - M_{x_j})}{S_{x_j}} & \text{if the indicator j has negative polarity} \end{cases}$$

where M_{x_j} e S_{x_j} are, respectively, mean and standard deviation of the indicator *j*.

The synthex index, for the unit *i*, is given by the formula:

$$MZ_i = \frac{\sum_{j=1}^{m} z_{ij}}{m}$$

Relative Indices Synthesis (MR)

Given the matrix **X**={x_{ij}} of *n* rows (units) and *m* column (indicators), the transformation matrix **R**={r_{ij}} is built, with:

$$r_{ij} = \begin{cases} \dfrac{(x_{ij} - \mathrm{Min}_{x_j})}{(\mathrm{Max}_{x_j} - \mathrm{Min}_{x_j})} & \text{if the indicator j has positive polarity;} \\ \dfrac{(\mathrm{Max}_{x_j} - x_{ij})}{(\mathrm{Max}_{x_j} - \mathrm{Min}_{x_j})} & \text{if the indicator j has negative polarity;} \end{cases}$$

where Min_{x_j} and Max_{x_j} are, respectively minimum and maximum of the indicators *j*.

The synthex index, for the unit *i*, is given by the formula:

$$\mathrm{MR}_i = \frac{\sum_{j=1}^{m} r_{ij}}{m} .$$

Mazziotta-Pareto index (MPI +/MPI -)

Given the matrix **X**={x_{ij}} of n rows (units) and *m* columns (indicators), transformation matrix **Z**={z_{ij}}, is built with:

$$z_{ij} = \begin{cases} 100 + \dfrac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} 10 & \text{if the indicator j has positive polarity;} \\ 100 - \dfrac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} 10 & \text{if the indicator j has negative polarity;} \end{cases}$$

where M_{x_j} e S_{x_j} are, respectively, mean and standard deviation of the indicator *j*.

The synthex index, for the unit *i*, is given by the formula [5] :

$$\mathrm{MPI}_i^{+/-} = \mathrm{M}_{z_i} \pm \mathrm{S}_{z_i} \mathrm{cv}_i$$

dove $\mathrm{M}_{z_i} = \dfrac{\sum_{j=1}^{m} z_{ij}}{m}$; $\mathrm{S}_{z_i} = \sqrt{\dfrac{\sum_{j=1}^{m} (z_{ij} - \mathrm{M}_{z_i})^2}{m}}$; $\mathrm{cv}_i = \dfrac{\mathrm{S}_{z_i}}{\mathrm{M}_{z_i}}$.

A3. The synthesis method

Table A1 contains a list of the main features of available methods, to select a suitable synthex index to the phenomenon studied.

The advantages of a synthesis index can be summarized in:

(a) one-dimensional measurement of a complex phenomenon, (b) easy interpretation in respect to a set of elementary indicators ("dashboard"), (c) simplification of data analysis (in particular, ordering of geographical units). Table A1 shows the main features of available method:

Table A1. Methods of indices synthesis features

Method of synthesis	Main features
Mean of standardised values (MZ)	Compensatory method It is based on the arithmetic mean of the z-scores. Mean 0 range between -3 and +3. Assumption: the indicators have the same variability.
Mean of relative indices (MR)	Compensatory method It is based on the arithmetic mean of the relative values. Range between 0 and 1. Does not exist a fixed mean.
Mazziotta-Pareto index (MPI^+/MPI^-)	Non-compensatory method It is based on an arithmetic mean penalized on the basis of the imbalance of values. Mean equal to 100, range between 70 and 130 Assumption: the indicators have the same variability. It is applicable to both positive events (MPI-) and negative (MPI +). It can be divided into two parts: "medium" (compensatory) effect and "penalty" effect (imbalance).

Bibliography

Aureli Cutillo E. (1996), *Lezioni di statistica sociale. Parte seconda, sintesi e graduatorie*, CISU, Roma.

Delvecchio F. (1995), *Scale di misura e indicatori sociali*, Cacucci editore, Bari.

Harbison F. H., Maruhnic J., Resnick J. R. (1970). *Quantitative Analyses of. Modernization and Development*, Princeton University Press, New Jersey.

Maggino F. (2008), *La misurazione dei fenomeni sociali attraverso indicatori statistici. Aspetti metodologici*, Università degli Studi di Firenze – Working paper, Firenze.

Mazziotta M., Pareto A. (2011), Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario, *Rivista di Statistica Ufficiale*, n. 1/2011, pp. 63-79.

Mazziotta M., Pareto A. (2015). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, DOI 10.1007/s11205-015-0998-2.

Mazziotta M., Pareto A. (2013). Methods for constructing composite indices: one for all or all for one?, Rivista Italiana di Economia Demografia e Statistica Volume LXVII n. 2 Aprile-Giugno 2013

OECD (2008), *Handbook on Constructing Composite Indicators. Methodology and user guide*, OECD Publications, Paris.

Silvio-Pomenta J. F. (1973), *Typological study using the Wroclaw Taxonomic Method (A study of regional disparities in Venezuela)*, SHC/WS/316, UNESCO.

---

[1] An simple indicator is computed data by means the ratio of a “raw” to a reference base (for example, “per capita income”).

[2] The polarity of an indicator is the sign of the relationship between the indicator and the phenomenon to be measured

[3] A simple indicator is considered “substitutable” if a deficit of an indicator can be compensated by a surplus in other.

[4] A simple indicator is considered “not substitutable” if a deficit of an indicator cannot be compensated by a surplus in other.

[5] If the composite index is ‘increasing’ or ‘positive’, i.e., increasing values of the index correspond to positive variations of the phenomenon (e.g., the socio-economic development), then MPI- is used. Vice versa, if the composite index is ‘decreasing’ or ‘negative’, i.e., increasing values of the index correspond to negative variations of the phenomenon (e.g., the poverty), then MPI+ is used.

APPENDIX B
Maps of main road arches from OpenStreetMap

Chart B1. Primary roads. Year 2017 (Open Street Map)

Chart B2. Secondary roads. Year 2017 (Open Street Map)

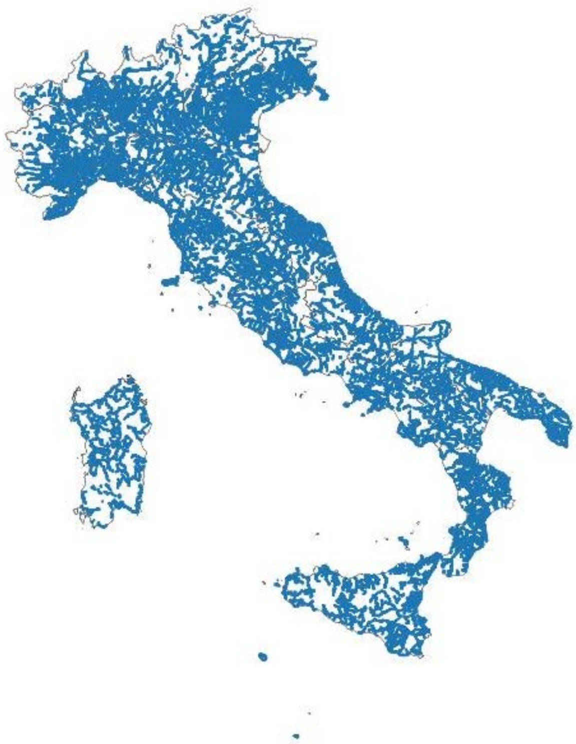


Chart B3. Motorways and trunks. Year 2017 (Open Street Map)

Chart B4. Residential living roads. Year 2017 (Open Street Map)

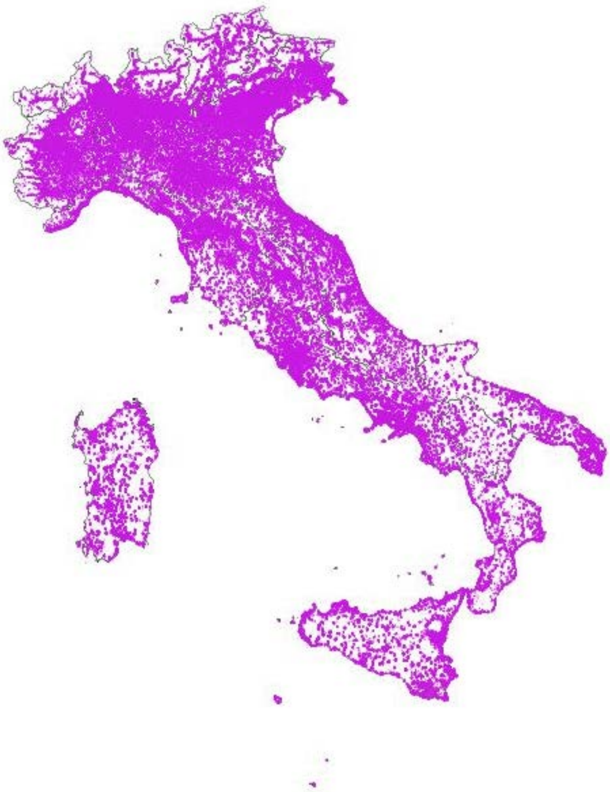


Chart B5. Tertiary roads. Year 2017 (Open Street Map)

Chart B1. Primary Roads. Year 2016 (Open Street Map)

Chart B2. Secondary Roads. Year 2016 (Open Street Map)

Chart B3. Motorways. Year 2016 (Open Street Map)

Chart B4. Trunks. Year 2016 (Open Street Map)

This experimental statistics has been produced by

- Silvia Bruzzone (bruzzone@istat.it; ph. +39 06 4673.7384); Marco Broccoli (broccoli@istat.it; ph. +39 06 4673.3312).