POLITECNICO DI TORINO
Repository ISTITUZIONALE

Road network comparison and matching techniques. a workflow proposal for the integration of traffic message channel and open source network datasets

Original
Road network comparison and matching techniques. a workflow proposal for the integration of traffic message channel and open source network datasets / Ajmar, A.; Arco, E.; Boccardo, P.. - In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. - ISSN 1682-1750. - ELETTRONICO. - 42:(2019), pp. 1503-1509. (Intervento presentato al convegno ISPRS Geospatial Week 2019, 10–14 June 2019, Enschede, The Netherlands.) [10.5194/isprs-archives-XLII-2-W13-1503-2019].

Availability:
This version is available at: 11583/2761232 since: 2022-06-12T17:58:29Z

Publisher:
International Society for Photogrammetry and Remote Sensing

Published
DOI:10.5194/isprs-archives-XLII-2-W13-1503-2019

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

ROAD NETWORK COMPARISON AND MATCHING TECHNIQUES. A WORKFLOW PROPOSAL FOR THE INTEGRATION OF TRAFFIC MESSAGE CHANNEL AND OPEN SOURCE NETWORK DATASETS.

A. Ajmar [1], E. Arco [2*], P. Boccardo [2]

[1] ITHACA, Via Pier Carlo Boggio 6, 10138 Torino, Italy - andrea.ajmar@ithaca.polito.it
[2] Politecnico di Torino – DIST, Torino, Italy – (emere.arco, piero.boccardo)@polito.it

KEY WORDS: Road Network, Data Matching, Conflation, Integration, Real-Time Traffic Data

ABSTRACT:

The rapid growth of methods and techniques to acquire geospatial data has led to a wide availability of overlapping geographic datasets with different characteristics. Road network data sources are today a significant number, with high differences in level of detail and modelling schemas, depending on the main purpose. In addition, continuous information about people and freight movement is today available also in real-time. This type of data is today exchanged between traffic operators using referencing standards as Traffic Message Channel. Integrating these heterogeneous databases, in order to build an added value product, is a serious task in geographical data management. The paper is focus on techniques to conflate the Traffic message Channel logical network on Open Source road network dataset, in order to allow the precise visualisation of traffic data also in real-time.
A first step of the research was the quality assessment of available Open Source (OS) road network dataset, then, a specific procedure to conflate data was set up, using an iterative process in order to reduce at every step the number of possible matching features. A first application of the enhanced OTM dataset is shown for the city of Turin: real-time open data of traffic flows recorded by road network fixed sensors, made available by the metropolitan Traffic Operation Centre (5T) and based on the TMC location referencing, are matched on the OTM road network, allowing a detailed real-time visualisation of traffic state.

1. INTRODUCTION

1.1 General context

In the transport sector, the ability to acquire continuous flow of data about people and freight movement has largely increased in recent years. In particular, cities are today full of traffic detectors, fixed and mobile, which monitor both private traffic and public transport movements. Raw measurements, as well processed data derived from these detectors, available in near real – time, can be classified as Geo Big Data (Ravanelli et al., 2018).
This vast amount of data is useful both for mobility management applications and for applications concerning territorial planning: the spatial representation of those data indeed allows to depict the behaviour of citizens in their daily movements, with respect to multiple temporal aggregations or time instants. This information can be integrated with other spatial elements in order to obtain new and interesting analysis of the society, which can be ingested as reference information to propose new urban planning policies.
In the context of the Project of National Interest (PRIN) URBAN-GEO BIG DATA (URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness), a goal is to setting-up an homogeneous road network dataset over the five Italian urban test areas (Turin, Milan, Padua, Rome, Naples), with a level of detail and characteristic suitable for a wide range of applications, as traffic monitoring visualisation, routing with real-time impedances, road asset inventory positioning. The results can be used at a later time for the implementation of a unique road network over the whole Italy (Ministero delle Infrastrutture e dei Trasporti, 2018).
In this paper, an application of open and big data of transport flow is described. Even if there is the availability of measures as open data, the use of those data in a spatial environment needs appropriate methodology to be exploited.
The area of study is the Turin City area, where the company 5T delivers a set of traffic flow measures as open data[1] (reference). These data are referenced using the Traffic Message Channel Standard, but in order to be corrected referenced over a more precise network, this location referencing must be converted on a more detailed road network.

2. METHODOLOGY

2.1 Road Network Datasets description

A first step of the research was the quality assessment of available Open Source (OS) road network dataset: a set of sources is described, focusing on their data structure, spatial quality and specific purpose. In particular, OpenStreetMap (OSM) road network, OpenTransportMap (OTM) data, Traffic Message Channel (TMC) network and the road network commercial dataset from HERE, available over Torino, were examined.

2.1.1 OpenStreetmap

OpenStreetMap (OSM) is a collaborative project with the aim to create and provide free geographic data, such as street maps, to anyone. The OpenStreetMap project arises as a solution of restriction of use and high cost of geographic data, positioning itself as a “cartographical Wikipedia”. Today OSM is considered a prominent example of volunteered geographic information, and, from the point of view spatial coverage and topographic elements, can be considered one of the best available sources of open data (Wikipedia contributors, 2018).

[1] http://opendata.5t.torino.it/get_fdt

The OpenStreetMap project stresses the difference between maps and geographic data: its main goal is to provide a world geographic database and not a map. A distinctive trait of OSM is that, instead of the layered approach used by most of geographic databases, it uses a single layer and a system of tags (a pair of key and value) for specifying the geographic meaning of a single object.
The data model is based on a tree data model, typical of XML structures. Four types of object can be defined: nodes, ways (ordered list of nodes), polygons (a way that starts and ends on the same node) and relations. Each object must have at least one tag and other tags can be added in order to describe the object: they can be view as attributes of the object and there are not strict rules on how many attributes any object should have (no upper limit). For transport purposes, the most common way to define a road is to edit an open way, an ordered list of nodes which also normally has at least one tag or is included within at relation; in case of road the tag "highway=*" is used. The value of the key "highway" is usually needed to indicate the importance of the road in the network, as "motorway", "primary", "secondary", "service" …. A list of possible values is specified in the Wiki of the project. Other tags can be combined, in order to define specific characteristic of the road ("maxspeed=*", "oneway=*" …). The definition of the direction of travel is of utmost importance for routing applications. "Forward" and "backward" describe a direction along a way, considering the direction in which the way is drawn in OpenStreetMap: "forward" means the digitized direction of the way, while "backward" means the opposite direction.
Looking at the use of OpenStreetMap data for transport purposes, the problem of data quality seriously limits its use: in particular, it is not always suitable for high-reliability routing applications. Limitations are related to a non-uniform spatial coverage in all areas. The high spatial inhomogeneity is also reflected at the attribute level. In in the case of a road sometimes only the geometric information is provided, the street names are sometimes included and sometimes not, and not always these names correspond to the official names, making difficult to exploit geocoding applications. Information about restrictions, number of lanes, maximum speed or direction of travel, as they are not mandatory, are often missing. Additionally, topology in network editing is frequently insufficient: roads are not always split at intersections, generating unconnected network and making OpenStreetMap not routable and not ready to use for analytical tasks.

2.1.2 OpenTransportMap

OpenTransportNet (OTN) is a European funded project focused on transport related services across Europe. A City Data Hubs has been constituted in order to aggregate transport-related data and spatial information, with the aim of creating innovative and collaborative ITS applications and services.
One of the first and most important outcome of the project is the OpenTransportMap (OTM) (http://opentransportmap.info/). The project aim was to provide a harmonized source of information for transport network, starting from OSM data, in order to enable routing services and expose data in the form of easy-to-understand visual interpretations, with a focus in the visualization of traffic volumes.
The INSPIRE Transport Networks data model was chosen as reference in order to build the harmonised data schema, as it addresses the linear topology and is compliant with the EU legislation. Processing steps have involved first the evaluation of geometry and in particular of the original topology of the OpenStreetMap data: some automatic and manual corrections have been applied over data in order to obtain a well-connected topology structure, which is critical to describe allowed movements between places (Jedlička et al., 2016).
Then the INSPIRE Transport Networks schema was analysed, selecting RoadLink and RoadNode as the main classes to be used for routing purposes. A mapping activity between OpenStreetMap tags and INSPIRE attributes and code lists has been made, with some information loss, when searching for the lowest common denominator: indeed, maybe some tags, which may be useful for roads classification, are not taken into account (Jedlička et al., 2016). Some limitations inherited from the OSM data still persist: attribute completeness is not optimal and non-homogeneous. The result is a road network enough suitable for routing and that can support the visualisation of traffic volumes.

2.1.3 Traffic Message Channel

The TMC indicates a standard service for spreading traffic information to final users through FM radio transmission and is mainly used by Traffic Information Centres to exchange traffic information. It is a logical network based on a set of related, ordered and well-known point location, identified by a unique identifier (Arco et al., 2017). Basically, the TMC standard defines the implementation schema of a TMC locations database, which represents the main road network implementing a sort of raw graph, with a set of points associated to a specific geographical point on the real road (e.g. main intersection, ramps, etc.) and arcs that connect couple of points, coded for machine-to-machine communication. Locations are the central element of the standard: they can be points on the road network, specific roads or part of roads, but also areas like municipality or other administrative units. Location codes are stored in location tables, together with additional information about the locations. In this perspective, a TMC Location Code (LCD), since is associated to a specific geographical location on the road network, it is used as reference for TMC events codification and localization. Thus transmitting only the numerical code all TICs and TCCs, as well as all mobile information services of a country, can understand exactly in which locations and on which way the event is located.
The TMC database for Italy (now at version 4.5) is hosted by the "Centro di Coordinamento Informazioni sulla Sicurezza Stradale" (CCISS), National TIC for Italy, which regularly update it and make it available on the Ministry of Infrastructures and Transport web site. It must be noted that TMC databases are not freely downloadable in each country: publishing it or making it available only to interested users is a national choice.

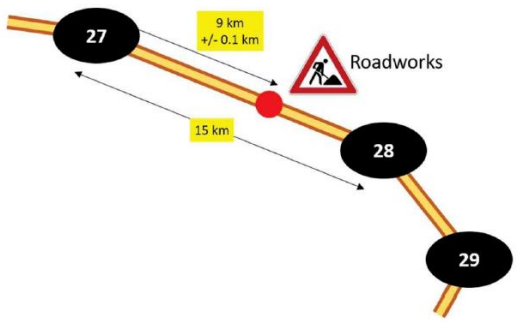

Figure 1: Georeferencing in TMC, readapted from Kamalski, 2005.

A TMC message consists of an event code, a localization code and some additional information such as the expiration time. The receiving system decodes messages and can display alerts on the map, via the TMC protocol and the localization tables already integrated in the maps of the major vendors. It can also present the message in text form, translating it into the user's language; all TMC receivers use the same list of event codes as the location database contains a set of country-specific location codes (Arco et al., 2017).
The localisation of the event is achieved through three main attributes: the primary location, where the event is located, the extent, which indicates other points involved, and the direction bit, which indicates the direction of the event-affected traffic flow. Linear referencing is applied in order to locate the event along a road, as can be seen in Figure 1.

2.1.4 NAVSTREETS Street Data

5T Agency in Turin uses NAVSTREETS Street Data, one of the main products of HERE, as main source of information for their operational activities.
The NAVSTREETS Street Data is delivered also in shapefile format (the one used in 5T) and provides, over the requested area, in addition to transport network data, other cartographic features that enhance the routing functionality. The delivered product contains several shapefiles and .dbf tables, but the main source of information is represented by the Streets shapefile, which are the navigable edges of the transport network, and the Zlevels shapefile, which represents nodes. Thanks to the definition of a specific set of rules for editing, the network is consistent over the whole area.
Looking into the schema of the Streets shapefile, the minimum mapping element is represented by the LINK, which have a minimum length of 2 metres and a maximum length of 10 kilometres.. LINKs can be grouped in FEATURE (LINK with the same "FEAT_ID" attribute) in order to defining roads with the same street name. The feature concept is useful from a semantic point of view, as it define as a unique element roads composed by several lanes or frontages, even if they are not continuous (the "FEAT_ID" is equal to zero in LINK representing complex intersection as roundabouts).
Each LINK is defined by a Reference Node ("REF_N_ID" attribute) and a Non-Reference Node ("NREF_N_ID" attribute), where first is the one with lower latitude or in case of the same latitude, with lower longitude.
According to the definition of Reference Node and Non-Reference Node, the navigability of the edge is defined as "BOTH", "FROM REF NODE" or "TO REF NODE" values in "DIR_TRAVEL" attribute.
The ZLevels shapefile contains both the necessary nodes for edges digitizing (internal nodes) and the nodes that define the intersections between links (identified by the "Y" value into the "INTRSECT" attribute). In addition, only intersection nodes provide a value in "NODE_ID" attribute (otherwise equal to zero), which allow to reference the link on which they belong (the value reported in "REF_N_ID" and "NREF_N_ID" in the Streets shapefile). Each node has the "LINK_ID" attribute that allows knowing which is the referenced edge.
The ZLevels shapefile contains the "Z_LEVEL" attribute that is used to represent junctions as the bridges and tunnels, so crossing over or under of links with other links. This attribute is not to be used to indicate actual elevation gain or loss, but to prevent routing between links that do not connect in reality ("0" for ground level, negative values for tunnel and positive for bridges).
The Streets shapefile contains over 98 attributes, which describes the characteristics of the road (Functional Class, Speed Category, Access …).
In addition, the NAVSTREETS Street Data set contain the Traffic table (.dbf), which enables the representation of Traffic Message Channel locations point over the road network. The information is stored at LINK level: for each element that has a correspondence with the TMC network a code is defined. Thanks to this reference, it is possible to locate traffic events along the network, having for each LINK the location code of the road and the location code of the point in positive and negative directions.
The NAVSTREETS Street Data, as an expensive commercial dataset, provides some assurances. A new version is provided each year, ensuring a high level of update. At the level of attributes, the data is almost complete and reliable, at least for the most important fields for routing (in general the level of attribute completeness varies among countries). It also has multiple names of streets and house numbers, allowing an advanced geocoding and routing applications.
Spatial accuracy is declared and varies between + or - 5 metres for the absolute positioning, and + or - 1 metre for relative positioning. In addition to the spatial coverage it is generally complete, in particular as regards the paved roads traversable by cars.

2.2 Comparison and assessment of road networks

The road networks described in previous section have been compared, from the point of view of spatial and attribute completeness. In general, most of the approaches to evaluate the completeness of spatial data over a certain area are achieved through a comparison with established references ones, for instance data produced by public administrations and institutions. In this case, the analysis was performed using the road network commercial dataset from HERE, as a benchmark to assess data quality. The commercial dataset is indeed set as an established reference, thanks to its reliability in terms of spatial and attribute completeness, which is declared and documented.
For the comparison over the Piedmont Area, firstly a selection of all features intersecting the regional area has been applied (for OSM, OTM and NAVSTREETS Street Data). This procedure has been favoured to a "clip" operation in order to maintain the integrity of features that cross the administrative borders. In addition, as NAVSTREETS Street Data contains mainly roads traversable by cars, in order to obtain a more comparable dataset, some selections have been performed. In particular, for NAVSTREETS Street Data, features classified as not paved and features that have access restrictions for cars, buses, taxis, motorcycles and trucks have been excluded from the analysis.
For OpenStreetMap data, following the GeoFabrik mapping documentation, features classified as bridleway, cycleway, footway, path, pedestrian and steps have been excluded, whereas no information about paved roads is reported in mapping schema.
For OpenTransportMap data, features classified as 'bicycleRoad', 'enclosedTrafficArea' and 'walkway' have been excluded. Even if the OpenTransportMap mapping schema reports the information about pavement, this information has been considered not reliable and consequently not used for selection. Indeed, the schema maps as unpaved all roads that not have this tag information, but it is more common (almost in Italy and Piedmont) to not have this tag on common roads than

the opposite. Then selecting unpaved roads, in this case, potentially excludes most of paved roads.
After these steps, it was evident that the OpenStreetMap data sources have more cars routable km compared to the NAVSTREETS Street Data one. This can be due to several reasons: OSM sources have in general more service streets, unpaved roads have not been taken into account, selection criteria applied cannot assure the correctness of data, also for eventual mapping errors inherited from the original OSM source. This discrepancy is particularly evident in mountain areas. In Table 1, the comparison of total km and number of features between OpenStreetMap, OpenTransportMap and NAVSTREETS Street Data is shown.

		NAVSTREETS	OSM	OTM
Piedmont - all	Total Km	58.924,85	101.324,40	128.300,77
	N° features	462.875	306.465	702.501
Piedmont - car traversable	Total Km	48.935,19	82.051,99	105.459,93
	N° features	406.291	255.984	609.952

Table 1. Comparison of total km and number of features between OSM, OTM and NAVSTREETS Street Data in Piedmont Region.

Following the approach described by Ludwig et al. in 2011, in order to have a more refined dataset to be compared with NAVSTREETS Street Data, features of OSM and OTM datasets within a distance of 5 m, 10 m and 30 m from NAVSTREETS Street Data have been further selected. This operation has been done in order to refine the matching between datasets, and to find a correspondence at object level. The desirable condition is to find a 1:1 correspondence between objects, which is obviously not possible.

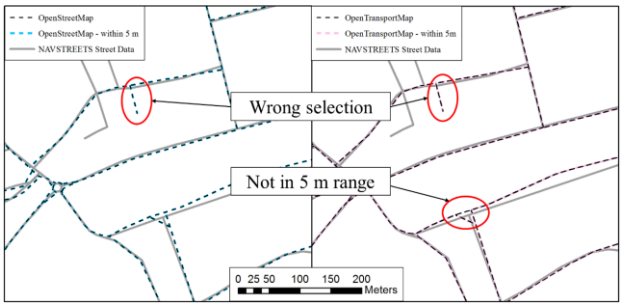

Figure 2: Selection errors applying the "within a distance of 5m" criteria.

The range of 5 m seems the most appropriate, as using larger ranges increase the spreading of matching errors: indeed, even if this operation reduces differences in datasets, results also in the selection of road branches that not corresponds logically to the original road from the NAVSTREETS Street Data, as can be view in Figure 2. Total Km and features count for selected datasets is shown in Table 2, from which can be observed how the 5 m buffer allows to have a better matching between OSM sources and NAVSTREETS Street Data.
Data comparison has been set also checking the distribution of Functional Class. In Table 3, the attribute matching between datasets used for the analysis is shown. This approach is prone to error, as it inferred many simplifications.
Main differences in the map matching at functional class level concern:

- ramps and junction classification, where in NAVSTREETS Street Data are usually classified in the lower class whereas in OSM sources in the higher one;
- service roads of major highway, where usually in OSM sources there is not a separation whereas are differently classified in NAVSTREETS Street Data;
- fourth and fifth classes in general.

	Total km	Features count
NAVSTREETS	48.935,19	406.291
OpenStreetMap	82.051,99	255.984
OpenStreetMap 5m	66.052,86	203.304
OpenStreetMap 10 m	67.454,18	210.398
OpenStreetMap 30 m	69.335,58	220.199
OpenTransportMap	105.459,93	609.952
OpenTransportMap 5 m	81.406,44	498.966
OpenTransportMap 10 m	83.259,50	514.430
OpenTransportMap 30 m	85.847,40	534.555

Table 2. Comparison of total km and number of features between OSM, OTM and NAVSTREETS Street Data in Piedmont Region, distributed for buffer selection.

The results are shown in Figure 3. In general, more similarities between OSM sources and NAVSTREETS Street Data can be found in the group of 5 m selection. OpenTransportMap also seems to have a more matching correspondence with respect to the GeoFabrik – OpenStreetMap source.

NAVSTREETS	OSM	OTM
1	motorway, motorway_link, trunk, trunk_link	mainRoad
2	primary, primary_link	firstClass
3	secondary, secondary_link	secondCLass
4	tertiary, tertiary link	thirdClass
5	all other values	fourthClass and fifthClass

Table 3. Attribute mapping between NAVSTREETS Street Data Functional Class and OSM and OTM datasets

Figure 3: Data comparison for functional class, as percentage of total km.

Another characteristic that can be assessed is the presence of an attribute name value, which is relevant for routing applications. In NAVSTREETS Street Data, as already stated, edges without name are the ones describing complex intersections. For this data source, features without a name are the 12% of the total. For OSM and OTM data, selection has been performed looking at two attributes: "name" and "ref" for GeoFabrik –

OpenStreetMap, and "roadname" and "nationalroad" for OpenTransportMap. If both attributes were not filled, the feature was considered without a name (without considering possible errors in name values). As can be view in Figure 4, features without a name are over the 50% for OSM, whereas percentages generally lower for OTM (nevertheless considerably high compared to NAVSTREETS Street Data). It is also evident from the figure that the buffer increase leads to select more features without name, probably because it has led to include more roads of the lowest hierarchy level.

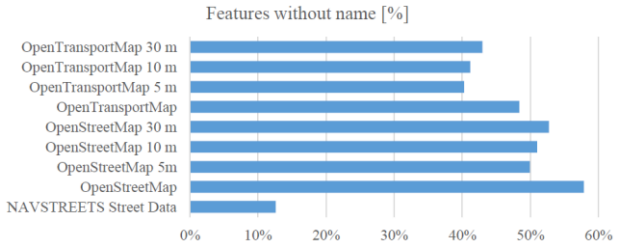

Figure 4: Comparison of number of features without a name, as percentage.

Finally, the topological correctness of data sources has been assessed. In particular, OpenStreetMap data are often not split at intersections: this is evident also in the higher value of mean length for a feature in OpenStreetMap, compared to the other two data sources, as highlighted in Table 4. A 'Feature To Line' operation has been performed in ArcGIS in order to evaluate the number of feature not split. New values can be found in the last row of Table 4: it can be stated that after this operation values between OSM and OTM are closer (as probably OTM have already these corrections implemented). It must be highlighted that the Feature to line operation can also generate some error for routing application if applied to the whole dataset: if it is true that connection must be guaranteed at intersections, this may not be correct in case of tunnels and bridges, where intersection between roads does not mean a real connection.

	NAV STREETS	OSM	OTM	OSM - Feature To Line
Feature Mean Length	0,127	0,331	0,183	0,170
N° of Features	406.291	255.984	609.952	481.524
Must Not Self-overlap	0	44	5	0
Must not Self-intersect	0	74	15	0

Table 4. Comparison of mean feature length and topology errors.

Other topological issues can be due to not connected roads and self-intersection. Evaluating through automatic procedures non-connected road is not trivial: for instance, the topological rule "Must not have dangles" in ArcGIS can be applied, but there is no way to discriminate errors from ends of road. An evaluation has been performed instead using the rules "Must Not self-overlap" and "Must not self-intersect". Results are listed in Table 4. Errors are higher in GeoFabrik – OpenStreetMap, even if the number allow a possible manual correction. As already stated, the biggest problem is related to un-connected roads and methods for assessment and correction must be implemented.

After this first analysis, OpenTransportMap has been chosen as source data set for the five Italian urban test areas. Transferring TMC information on the more detailed OTM network will allow a precise localisation and visualisation of traffic events, also in real-time.

2.3 Conflation procedure of TMC to OTM

As the TMC network is only logical (connection between points do not represent the real path of the road), techniques to match this data with a reference network must be applied in order to properly visualise measures.

In order to achieve this goal, a specific procedure to conflate data was set up, using an iterative process in order to reduce at every step the number of possible matching features. The procedure has been implemented using PostGIS functions and PostGIS Routing extension.

The main steps of the procedure are the following:

1. A matching between OTM nodes and TMC nodes is performed;
2. For each TMC road, a routing algorithm is performed on OTM network, searching for the minimum path between two nodes, for each direction of travel (positive and negative).

The selection of OTM crossing nodes matching specific TMC points was performed first using proximity criteria. For each TMC nodes a set of maximum 5 OTM crossing node have been selected. Note that only OTM nodes which a degree > 2 have been considered selectable (crossing node).

This selection was refined through the analysis of road names. For each crossing node of OTM, a list of road names related with it is associated. TMC points instead have already a road name associated (). Using Levenshtein distance, the previous set of selected OTM nodes is compared with TMC points names. After this step, a sensible reduction of the OTM nodes selected is obtained. Mismatching still persist in case of double digitized roads, where is not possible in an automatic way to know which a TMC point pertain to a specific direction of the roads. This new set of point is characterised by the TMC code, the ID of the (one or more) OTM point associated and the geometry of the OTM point.

The second step has been performed writing a specific function in PostGIS, which iterates over the TMC roads. For each road, it looks at the sequential list (first for positive direction, then for negative one) of TMC points which compose the road. For each couple of consecutive points, the algorithm uses the geometry derived from the set previously created and use the Dijkstra Shortest Path in order to find the minimum route connecting the two points. The Dijkstra Shortest Path has been customised in order to search only for roads which are at least in the "thirdClass" hierarchy, as usually TMC roads are only main roads. This condition has been removed only for those couples of points where no solution was found.

As a TMC points may be related to one or more OTM point, multiple solutions have been found. The final step of the function selects the minimum length of the paths for each multiple solution founded.

The proposed solution allows to partially automate the conflation of TMC in the OTM datasets, even if is still prone to errors and highly related on the specific used datasets. In particular, there is the need of an interactive checking of the results between steps.

The resultant paths were associated to the OTM network, adding an attribute to OTM edges and nodes, identifying TMC roads, start and end TMC point for negative direction and start and end point for positive direction.

After these two steps, new attributes in the OTM data allow the representation of the TMC network. Over this new network,

locate measures along monitored arcs can be easily achieved through linear referencing techniques.

2.4 Traffic flow open data description

In this section will be given a description of the sensors involved and of the data structure used to deliver the information, deepening the S.I.MO.NE. protocol, an XML schema designed by 5T to share traffic information. S.I.MO.NE., acronym for "Sistema Innovativo di gestione della MObilità per le aree metropolitaNE" (Innovative System for Metropolitan Area Mobility management), is an Italian project, coordinated by 5T company, devoted to implement a Decision Support Systems (DSS) and standard communication protocol to address private mobility management (Arneodo et al., 2009).
In particular, the S.I.MO.NE protocol takes into account two main spatial reference systems in order to locate properly the information over the road network, the WGS84 coordinates and the TMC, which are of primary interest in order to allow the correct visualisation of traffic measures.
The project has developed a communication protocol to exchange information from vehicles fleets to Fleet Manager Centres and different TCCs in a bidirectional way, and a software component which aggregates measures. This last aspect, which regards the definition of a level of accuracy of the traffic estimates, respect to the quantity of vehicles in the traffic, and methods to aggregate and integrate measures, will be not deepened in this research. What is of interest here is the structure of the XML Schema realised to exchange those information, as S.I.MO.NE. is considered an Italian Standard, already adopted by several cities as Bologna, Genoa, Florence, Cagliari and Turin (which was the test sites of the project) and have the potential to be spread across the country.
Three XSD (XML Schema) have been developed (freely downloadable from S.I.MO.NE. project site): one for managing traffic data gathered from FCD and other sensors, and following aggregation as historical profiling (traffic monitoring and control), one for managing traffic events (info mobility) and last for managing Restricted Access Area (RAA). The first schema is the one of interest in this context.
In particular, the traffic data schema defines a root element traffic_data, as a container for all traffic data types: it has attributes defining the type of data contained, the time generation of the data, the start and end validity time for data and the identifier of the data provider.
Through an HTTP request, the agency deliver as open data the measures of speed and flow detected by a set of approximately hundred traffic sensors, updated every five minutes. This set is delivered in XML format and need appropriate transformation to be used in a GIS environment.
In particular, the location referencing through TMC is achieved giving: the TMC road code and name, the TMC point to be used as start point, an offset distance (distance calculated over the road from the TMC point) and a direction of travel (use to decide how to apply the offset distance). These attributes have allowed, through a simple PostGIS function, to precisely locates traffic detector over the OTM road.

3. RESULTS

3.1 Implementation of the visualisation platform

A first application of the enhanced OTM dataset is shown for the city of Turin: real-time open data of traffic flows recorded by road network fixed sensors, made available by the metropolitan Traffic Operation Center (5T) and based on the TMC location referencing, are matched on the new OTM road network, allowing a detailed real-time visualisation of traffic state.
Firstly a custom function in POstgreSQL has been written. This function, thanks to the HTTP extension, and the use of PgAgent, is triggered every 5 minutes, and make an HTTP request to the open data URL exposed by 5T. The XML returned is parsed in order to be stored on a table: each properties of the XML are mapped into specific fields. Other functions take care of making daily and monthly backups.

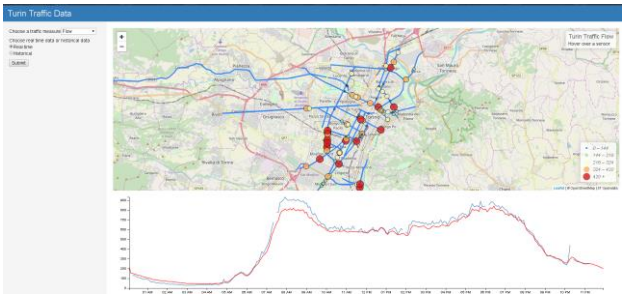

Figure 5. Overview of the web map.

With this framework, a daily table store all the incoming data, from the 00:00 of the day to the last five minutes. Others table historicize the data month for month, preserving only the ID of the measurement station, the start and end time to which the measures are referred and the values of flow, speed and accuracy. The amount of historicized row is now of 8'481'857, and time range spans from 12 April 2018 to now.

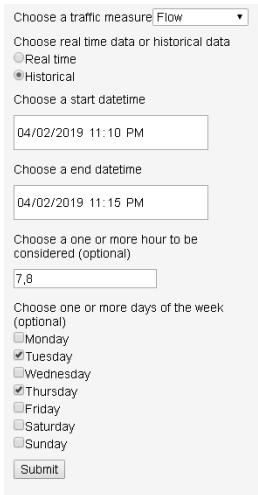

Figure 6. The expand form for selecting historicised data.

On the other side, a web map has been built to allow the visualisation of historical and near real-time traffic flow and speed over the Turin main road network. The web map has been built with NodeJS and Express framework, using LeafletJS for geographic data visualisation.
As default, the near real time flows are displayed (Figure 5). A form on the page allow to choose other measures (speeds) and time ranges. Each request from the platform is managed with specific PostgreSQL functions with specific parameters: in particular is possible to choose a specific temporal window, and within the window is possible to select specific hours (as instance only peak hours) and/or specific week days. In general, a filter is applied in the query, selecting only data with accuracy value major of 70. An overview of the form is visible in Figure 6.

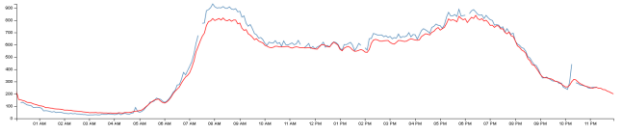

Figure 7. The graph showing the current day traffic profile (blue) and the mean traffic profile of the weekday (in red).

At now the web map integrates also a graph (made with D3js) which shows the traffic profile of the current day and mean traffic profile of the weekday extracted from the historicised data (Figure 7), helping in the understanding of the general dynamic of the traffic flow.

4. CONCLUSION

The conflation algorithm developed for this research can be applied may be also applied in other areas and potentially over the whole Italian network. This objective is one the goal listed in the Italian Smart Road Initiative, where the use of TMC is encouraged between TOCs. This matching indeed can be useful for other areas where open transport data are shared with a TMC reference (as for motorways traffic event information). In addition, it can be a first step to overcome the use of commercial road network on cars navigation systems.
Further developments will consider the possibility to locate also a set of traffic events with TMC location (not available at now).
On the other hand, the visualisation of traffic flow, made available as open data in fully open source environment, enables the integration of these data with other spatial information, creating new services, analysis and spatial applications. As instance, the traffic flow detected during particular social event can be linked with social network, enabling the discovering of new information about people behaviours and preferences. Public Sector in particular, can benefits from linking and visualising different data sets and be able to make better public service decisions based on the findings. In addition, as data gathered are continuously increase, new strategies for storing and especially querying data in an efficient way, must be implemented.
Finally some improvements on the web map will be take on, like enriching the possibility of graph visualisation in order to better understanding the aggregated data and the possibility to infer the traffic state level (over roads stretches) starting from the point data already available.

ACKNOWLEDGEMENTS

URBAN-GEO BIG DATA (URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness) is a Project of National Interest (PRIN) funded by the Italian Ministry of Education, University and Research (MIUR) – id. 20159CNLW8.

REFERENCES

Arco E., Ajmar A., Arneodo F., Boccardo P., 2017. An operational framework to integrate traffic message channel (TMC) in emergency mapping services (EMS). In: *European Journal of Remote Sensing*, vol. 50-1, pp. 478-495. ISSN 2279-7254. DOI: 10.1080/22797254.2017.1361306.

Arneodo, F., Foti, G., Cocozza, M., 2009. S.I.MO.NE. Innovative system for metropolitan area mobility management. In: *Proceedings of the 16th ITS World Congress*, 21-25 September 2009, Stockholm, Sweden.

Jedlička K., Hájek, P., Cada, V., Martolos, J., Šťastný, J., Beran, D., Kolovský, F. and Kozhukh D., 2016. *Open transport map — Routable OpenStreetMap*. In: IST-Africa 2016 Conference Proceedings, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information Management Corporation. Durban, South Africa, 11 - 13 May, 2016. ISBN: 978-1-905824-54-0

Kamalski, T., 2005. European experience with TMC: Supplying value. *TMC workshop*, Brussels.

Ministero delle Infrastrutture e dei Trasporti, 2018. Modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni di Smart Road e di guida connessa e automatica. *Gazzetta Ufficiale della Repubblica Italiana, Serie generale*, n.90. Available from: https://www.gazzettaufficiale.it/eli/id/2018/04/18/18A02619/sg (30 March 2019).

Ravanelli, R., Crespi, M., 2018. Analysis of the floating car data of Turin public transportation system: first results. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-4. 515-521. 10.5194/isprs-archives-XLII-4-515-2018.

Wikipedia contributors, 2018. OpenStreetMap. *Wikipedia, The Free Encyclopedia*. Available from: https://en.wikipedia.org/w/index.php?title=OpenStreetMap&oldid=824772218 (20 February 2018)

Revised April 2019