View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by idUS. Depósito de Investigación Universidad de Sevilla

Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Procedia 18 (2016) 320 – 327

XII Conference on Transport Engineering, CIT 2016, 7-9 June 2016, Valencia, Spain

Harmonization of transport data sources according to INSPIRE data specification on transport networks

Cayetano Ruiz-Alarcon-Quintero[a]*

[a]*Transportation Engineering, ETSI - Universidad de Sevilla, Camino de los Descubrimientos s/n, ,Seville 41092, Spain*

Abstract

There are numerous sources of data referred to transport from different suppliers, the main one being the Public Administration. According to the objective of each project, datasets are collected and formatted in a particular way. Each public organization only covers the area that corresponds to it (municipal, provincial, regional, national), and this results in the network being covered in a complementary manner, but with no interoperable datasets formats. Conduct studies and projects whose area of study exceeds more than one scope involve making a first phase of harmonization of data with high costs. These costs could be reduced by using standard formats and interoperable data sets and services. The European Union has stimulated the process of standardization and interoperability through INSPIRE Directive (2007/2/EC), which will be fully implemented in 2019 in the Member States. In this paper a harmonized data model of existing datasets provided by the traffic detectors in urban areas of Seville and Malaga is proposed, because they are used in many traffic projects and studies. The main objective is subsequent adaptation to the proposal by the INSPIRE directive specification, taking into account compatibility with the standard format DATEX II. This process can help to accelerate the transition to the standardized format to which the European Union is directing transport networks, and data sets associated with them.
© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the organizing committee of CIT 2016

Keywords: Standardization; Harmonization; INSPIRE; Interoperability; Traffic detector; Traffic flow

1. Introduction

One of the first phases of transport engineering projects is the collection of existing data on the field of study and the area in which of execution, if the project has a specific geographical area.

To cover this phase of the project it is necessary to search for the available series of data, both public and private. In studies related to the modeling of transport, mobility surveys are often used, covering different areas of transportation, demographics, traffic measurements detectors, network topology transport and different features. The

* Corresponding author.
E-mail address: cruizdealarcon@us.es

2352-1465 © 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the organizing committee of CIT 2016
doi:10.1016/j.trpro.2016.12.043

data presented may have geographic format, such as the location of traffic detectors, or topology of the transport network; but no geographic data as intensity or the capacity of a section of the network is also required, and must be related to geographic entities like a point or a link of the network.

The costs of this phase could be reduced if the data were produced in a standardized format, taking into account the timing of the records and publishing them, enhancing their use and interoperability through open licenses, European Union (2015).

The main motivation of this paper is the experience accumulated in this department over various projects whose main objective was to establish a transport demand model. Most of the time has been consumed in the compilation and harmonization of data. During the development of these projects various proposals have emerged from the team involved. It is intended to adapt the proposed solutions to the framework provided by INSPIRE and the thrust of implementation in European cities and regions policy of open data and open government.

INSPIRE European Parliament (2007), arises from the lack of availability, quality, organization, accessibility, and sharing of spatial information are common to a large number of policies and activities, and are experienced across the various levels of public authority in Europe. In parallel with this directive the European Commission is promoting the use of standards, particularly in the field of transport data collection, the proposed EU standard is DATEX II, Commission Delegated Regulation (EU) No 962/2015. Sheltered by this policy have been launched pilot projects to test the benefits and potential problems of this new framework such us Open Transport Net, Humboldt, Plan4all and Plan4Business projects, also in the field of standardization and interoperability related to the world of transport. There are also non-institutional solutions in this field as OSM, which is an interoperable network throughout the world, under the OGC.

A process of harmonization of data, focused on traffic detectors (flow, speed, occupancy, etc.) from different sources in the same geographical context, urban areas of Seville and Malaga is proposed.

It is intended to verify the potential of the engineering projects and research data published in open data and accessible formats, which is in line with the philosophy of the INSPIRE directive.

Nomenclature

AAWDT	Annual Average Weekday Traffic
AADT	Annual Average Daily Traffic
AAHWDT	Annual Average Hourly Weekday Traffic
MADT	Monthly Average Daily Traffic
K_{30}	The proportion of Annual Average Daily Traffic (AADT) occurring during the 30th highest hour of the design year.
K_{50}	The proportion of Annual Average Daily Traffic (AADT) occurring during the 50th highest hour of the design year.
K_{100}	The proportion of Annual Average Daily Traffic (AADT) occurring during the 100th highest hour of the design year.
OGC	Open Geospatial Consortium
OSM	Open Street Map
INSPIRE	European directive establishing an infrastructure for spatial information in Europe
LISIGE	Spanish law that ensures the implementation of INSPIRE directive in Spain
IGN	National Geographic Institute of Spain
IGR	Geospatial Information Databases of Reference in Spain
TWG	Thematic Working Group of INSPIRE
RT	IGN Transport Network
OTN	Open Transport Net project
DATEXII	standard for information services traffic in real time
CCT	Center Mobility Management
SDI	Spatial Data Infrastructure
PostGIS	Postgre complement for GIS data
PgSQL	Postgre SQL language

GIS	Geographic Information Systems

2. Harmonization of data transport, European regulatory framework.

2.1. ITS Directive and DATEXII

Commission Delegated Regulation (EU) No 962/2015 and ITS directive, European Parliament (2010), advises States stick to the DATEX II standard for information services traffic in real time. In addition to this regulation states that the road authorities and infrastructure operators provide dynamic data of road conditions (art. 5) and traffic data (art. 6) in that format, that regulation will be mandatory from July 13, 2017.

2.2. INSPIRE Directive

INSPIRE aims to establish a SDI, focused on environmental policies and activities that impact the environment. The philosophy of INSPIRE is to tap the resources and data already available in the Member States, trying to raise the minimum possible new data, to have a unified data format from the beginning, European Parliament (2007).

The INSPIRE directive came into force on 15 May 2007 and will be implemented in various stages, with full implementation required by 2019.

Specifically for the field in question, this determines a particular specification for transport networks, INSPIRE TWG Transport Networks (2014). This specification document must be followed by the States Members Authorities in the process of collection and storage of data.

It should include an integrated transport network and related elements, which are continuous in each national limit. The topographical features of the transport network are related to road, rail, water transport and air. There should be links between different modes. The transport network should support the reference to transport flows to enable navigation services. The latter is one of the most important elements in this study, as the ultimate target is to have a homogeneous format for data traffic detectors from different sources, which, among other data, the traffic flow.

Spain has just startedto adapt existing data regarding transport network to this specification, Grupo de trabajo de Redes e Infraestructuras del Transporte (2015), in production by the IGN. This network aims to meet the specifications of the new Plan Production IGN, which are aligned with the specifications listed in Annexes I and II of INSPIRE and Annex I of the national law that transposes law LISIGE, Ley 14/2010.

The INSPIRE specification provides an option to store data collected by a sensor INSPIRE Cross TWG on O & M (2013) considering they are specific elements of the type "Features of Interest Point" with "Multiple Results in Time". This proposed implementation might be appropriate for the values that collect traffic detectors, such as intensity, occupation or the speed detected at a certain time and which takes into account the temporal aggregation and comparison of data. However, it should be noted that this issue relates to sensors that collect environmental data, so it may not be entirely appropriate from the point of view of interoperability, since it is not clearly specified that this specification is used for flow of traffic or traffic conditions in real time.

3. Harmonization of traffic flow data in the urban areas of Seville and Malaga

All problems arising from heterogeneous data sources are observed in this case. Not harmonized, and not interoperable, and in many cases datasets are not directly machine-readable as dictated by INSPIRE. A brief analysis of the two urban areas and a description of the detectors in them are done.

3.1. Area of study

The geographical context in which the study is developed are the urban areas of Seville and Malaga. Urban areas are local entities integrated by the municipalities of large urban agglomerations population centers whose economic and social links demand planning and coordination of certain services and make necessary works together, Ley 7/1985.

The urban area of Seville is composed of 24 municipalities, Areas Urbanas +50 (2012). Main municipalities are Sevilla, Dos Hermanas, Hermanas, Alcalá de Guadaíra, Mairena del Aljarafe, Coria del Río and Los Palacios y Villafranca. In 2012 this urban area housed a total of 1,294,867 inhabitants, of which approximately 700,000 reside in Seville, another great concentration to the southeast, in Dos Hermanas and Alcalá de Guadaira with 200.000 inhabitants, and another large concentration of inhabitants in the Aljarafe, west of the capital.

The urban area of Malaga includes 8 municipalities with a total of 953,251 inhabitants, Areas Urbanas +50 (2012). These are Malaga, Mijas, Fuengirola, Torremolinos, Benalmadena, Rincon de la Victoria, Alhaurin de la Torre and Cártama, in which the core is Malaga 568,479 inhabitants.The detectors in the Urban Area of Málaga are distributed similarly to Seville.

In the ring there are DGT and Ministry of Development detectors, detectors in intermunicipal routes belongs to Government of Andalusia, and within the urban area detectors belongs to council.

3.2. Suppliers, format and location of detectors

In both areas there are 5 different administrations collecting datasets about traffic flow. Even the Government of Spain itself there uses two different sources of detectors, Mapas de Tráfico (2014) and Mapa del Tráfico (2016).

In both urban areas we find the following detectors according to the ownership of the road. Notably, most of the data are published in non-machine-readable formats, which greatly hinders its extraction and use for the user.

Table 1. Traffic flow data suppliers.

Traffic flow data suppliers	Parameters	Detail of data	Detectors
Traffic maps, Ministry of Development	MADT	Road	101
Map Traffic, Traffic Department, Ministry of Interior, Government of Spain	Records each 3 minutes on-line (Flow, occupation, velocity, light vehicle percentage)	Road and direction	166
Gauging, Ministry of Public Works and Housing, Government of Andalusia	AAWT AAHWT K_{30} K_{50} K_{100}	Road and direction	87
CCT Sevilla Center Mobility Management, Delegation Security, Mobility and major festivals - City of Sevilla	AAWT AAHWT Records each 5 minutes on-line (Flow, occupation)	Road and direction(annual summary) Road, direction and lane (on-line)	49 (annual summary) ~ 180 (on-line)
Government area of accessibility and mobility, Malaga City Council	PHF AADT AAWT	Road and direction(quarter summary)	49

4. Adaptation to INSPIRE specification

4.1. Harmonization of traffic flow detectors data

The criterion followed is to collect as detailed as possible data, as recommended in the INSPIRE Directive, European Parliament (2007) and the Working Group on Open Data, European Union (2015). Later, we need to make all data interoperable even if they are obtained from a different source. This requires taking into account the frequency with which the data is recorded, how often it is published, and a number of features inherent in the data that may change depending on the source.

The choice has been to design a data model that allows to store all this information, with a quick automatic processing, in an interoperable format. Once an absolute interoperability is achieved, the adaptation to INSPIRE and RT is greatly facilitated.

4.2. Traffic flow in INSPIRE specification

In INSPIRE TWG Transport Networks (2014), the data collection of vehicle flows is cited at the beginning of the document. The transport network should support the reference to transport flows to enable navigation services.

Although the objective of using the network for navigation, this specification is allows the storage and representation the traffic data. In cases of use of application of this specification flow modeling, capacity planning, information systems in vehicles, travel planning, navigation, routing, traffic control and traffic management is contemplated.

In the data model the direction of traffic flow is taken into account within the "Transport Property" entity, particularly in the "Traffic flow Direction" entity. However, it is observed that is not any field in which the flow value in an arc or a lane is stored. This is understandable because it is giving a static geographical context, which are considered the physical and geometric characteristics of the transport network and the elements that complement (such as traffic signals or stations), but no variables dynamics in an immediate timeframe.

4.3. Data model of harmonized data

To collect data from different sources is proposed the following data model to be stored in a PostGIS database, which will have major entities, and entities associated with them to complete the information. PgSQL scripts are used in order to transform data from different sources and add them to the database, Pedro O. Santos et al. (2015).

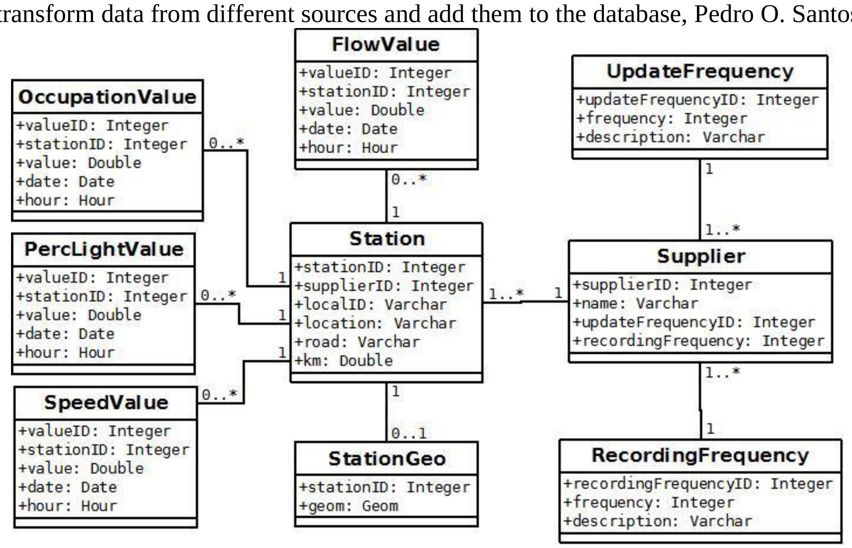

Fig. 1. Proposed data model.

In the data model proposed for detectors, Fig1., it has been respected the granularity with which the data is published originally. Information about the frequency of recording and publishing data is necessary for comparison between different sources, which is useful for prospective studies and analysis.

4.4. Connections of data model to OSM, INSPIRE and RT networks.

The purpose of this document is to link information from traffic detectors with existing standard formats or those to be set up in the near future in the European Union.

Once the data is harmonized, we must establish a connection point with the networks concerned. In Fig2. is shown the logical point of connection with networks. It has been considered in the data model that data detectors could be linked to different networks, which is why you can easily link with the various cases discussed. The data model detectors behave as a black box and directly contribute data flow on the arc.

The OSM has been chosen network to represent the transport network because the European Commission has established this network as a basis for projects to have their own interoperable network. some projects that use it are paln4business or OTN, Jèzek J., Jedlicka K., Martolos J. (2015). Using this network may allow a smooth transition to INSPIRE format when official networks under this specification begin running throughout Europe.

OSM is interoperable throughout the European territory and a routable network, with minimum acceptable precision, and with a data model that allows the calculations required for modeling transport on the network and is easily exportable, as it is published in open format Jedlicka K., Jezek J., Mildorf T. (2015) propose an adaptation of OSM data model to INSPIRE in the OTN project.

Specifically, the adaptation consists on a table of the database that would establish a relation between the station and the link of the network, Fig2. Each network has an identifier that allows to establish relationships with different networks from the stations.

This model is sufficiently versatile and flexible to link information produced at different sources with standardized networks proposed.

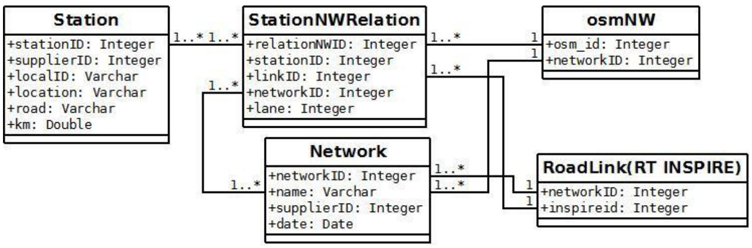

Fig. 2. Connection to different networks.

4.5. Case of use. Clustering of traffic flow detectors in Seville.

In this section, a case of use is considered; the clustering of the traffic flow detectors in the urban area of Seville with the aim of testing the database. This case considers the detectors supplied from Estado del Tráfico (2016) and Mapa del Tráfico (2016). Only detectors that provide real time data are considered. These data are recorded into the database automatically from the web.

R statistical and graphic computer software is used to apply the clustering, *PAM* (partitioning around medoids) is the method selected. It is similar to the well-known *kmeans* algorithm, Rousseeuw et al. (2015), but it is more robust because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances. The Calinski-Harabastz index Christian Hennig (2015), is used to determine the best k, resulting in the selection of k=3.

Flow pattern is normalized because detectors refer to different elements, the ones of Estado del Tráfico (2016) refer to urban lanes, and the ones of Mapa del Tráfico (2016) without disaggregation by lanes. This is a problem that remains: although the data is in the same format, it must be treated with care.

Detectors are separated basically into three groups. One of them, cluster 0, is composed by defective detectors. The other two have a complementary weekday traffic flow pattern. Cluster 1 is shown in blue, and cluster 2 in red in Fig3. In the same road but different directions, detectors located on ways that get into the main city belongs to cluster 1, and detectors in the opposite direction belongs to cluster 2.

If we eliminated the defective detectors, surely we would have a bigger k. A bigger number of clusters could show a more detailed road behavior.

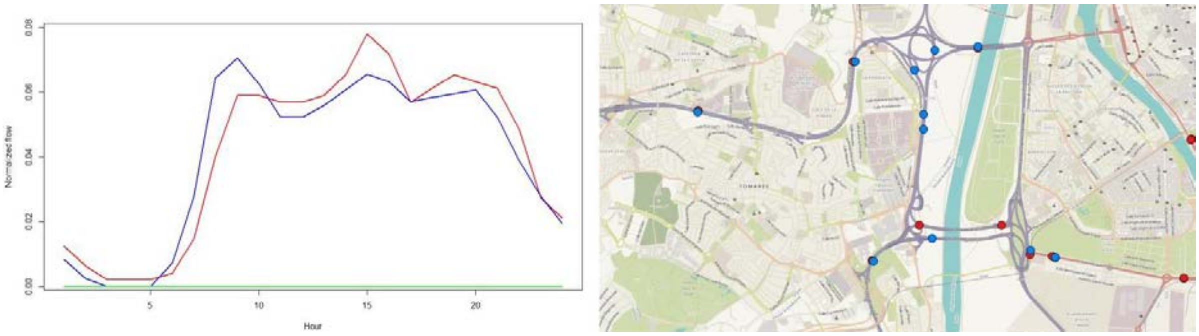

Fig. 3. Clustering medoids.

5. Conclusions

An intensive process of data harmonization of public administration is necessary and essential to extract the maximum potential to existing data sets. Interoperability of data and services by standardizing formats, from the most basic levels of government to the highest EU allows savings that have a strong economic impact. If this is accompanied by free licenses and accessibility for all citizens, it favors a more open society and a policy shared with citizens.

Harmonized data sets greatly facilitate the data collection phase in traffic engineering projects. When there is no such harmonization, this phase becomes a tedious, costly and in some cases impossible task. This affects the final quality of the projects and research.

In the case of data sets of detectors traffic in urban areas of Sevilla and Malaga, the status of harmonization, standardization and open data is in a very primitive state. It is necessary a great effort of technicians to take advantage of existing sources. With the database proposed in this paper, it has been proven the ease with which it can be crossed datasets from different levels of public administration, harmonization provides a rich data source that hasn’t been used until now. If these datasets were available for all potential users, it would be a powerful benchmark to prove test theories and studies. This is what is expected when policies promoted by the European Union will be implanted in 2019.

Projects such as OTN, or those related to Smart City being promoted by numerous local administrations, demonstrate the social and economic potential of standardization of data sets and services and accessibility to them.

That is why the implementation of standards such as INSPIRE and DATEX II accompanied by the promotion of policies that encourage the establishment of open licenses for public data will have a very positive way on the European Union.

Acknowledgements

This research has been possible thanks to financing the project “Advanced modeling techniques travel demand. Application to strategic and operational level in Spain” (TRA2012-36930) under the State Plan of Scientific and Technical Research Innovation 2013-2016, the Ministry of Economy and Competitiveness of the Government of Spain.

References

INSPIRE Cross TWG on O & M (2013). D2.9 Draft Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development

INSPIRE TWG Transport Networks (2014). D2.8.I.7 Data Specification on Transport Networks.

Grupo de trabajo de Redes e Infraestructuras del Transporte (2015). Especificaciones del producto Redes e Infraestructuras del Transporte

European Commission (2010). European Interoperability Framework (EIF) for European public services.

European Parliament (2007). DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

European Parliament (2010), DIRECTIVE 2010/40/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 July 2010

European Union (2015). Creating value through Open Data: Study on the Impact of Re-use of Public Data Resources.

Commission Delegated Regulation (EU) No 962/2015 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services

Jèzek J., Jedlicka K., Martolos J. (2015). Visual Analytics of Traffic-Related Open Data and VGI. ICIST 2015.

Jedlicka K., Jezek J., Mildorf T. (2015) D4.4 Data Harmonisation and Integration, OpenTransportNet - Spatially Referenced Data Hubs for Innovation in the Transport Section. Version 1.0 - 02/02/2015.

Rousseeuw et al. (2015) Finding Groups in data: Cluster Analysis.

Christian Hennig (2015). Flexible Procedures for Clustering.

Ley 7/1985, de 2 de abril, reguladora de las Bases del Régimen Local. Disposiciones Generales. Jefatura del Estado. Ref: BOE-A-1985-5392.

Ley 14/2010, de 5 de julio, sobre las infraestructuras y los servicios de información geográfica en España.

PedroO. Santos et al. (2015). Comparative performance evaluation of relational and noSQL databases for spatial and mobile applications.

Estado del Tráfico (2016). Portal de Datos Abiertos. Ayuntamiento de Sevilla. http://datosabiertos.sevilla.org/dataset/?id=estado-del-trafico

Intensidades de tráfico (2015). Área de Gobierno de Accesibilidad y Movilidad, Área de Movilidad. Ayuntamiento de Málaga. http://movilidad.malaga.eu/portal/menu/seccion_0008/secciones/subSeccion_0002

Areas Urbanas +50 (2012). Información estadística de las grandes áreas urbanas de España 2012. Ministry of Development. Government of Spain.

Mapas de Tráfico (2014), Ministry of Development. Government of Spain.

Mapa del Tráfico (2016), Dirección General de Tráfico, Ministry of Interior, Government of Spain. http://www.dgt.es/es/el-trafico/

Aforos (2014). Consejería de Fomento y Vivienda, Junta de Andalucía. http://www.juntadeandalucia.es/fomentoyvivienda/portal-web/web/areas/carreteras/aforos

Intensidades Medias Diarias de Vehículos (2015). Centro de Gestión de Movilidad, Delegación de Seguridad, Movilidad y Fiestas Mayores, Ayuntamiento de Sevilla. http://trafico.sevilla.org/imd.html

.