DELIVERABLE

Project Acronym: **OTN**
Grant Agreement number: **620533**
Project Full Title: OpenTransportNet – Spatially Referenced Data Hubs for Innovation in the Transport Section

---

D4.4 DATA HARMONISATION AND INTEGRATION

Version: 1.0

Authors:
Karel Jedlička (University of West Bohemia)
Jan Ježek (University of West Bohemia)
Tomáš Mildorf (University of West Bohemia)

Reviewers:
Lieven Raes (CORVE)
Pieter Colpaert (iMinds)
Jan Martolos (EDIP)
Dzintars Vilnis (EX)
Phil Archer (W3C)
Andrew Stott (CORVE)
Bart de Lathouwer (OGC)

Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

This project has received funding from the European Union's Competitiveness and Innovation Framework Programme under grant agreement no. 620533.

Table of Contents

List of Tables

List of Figures

List of Acronyms

API	application programming interface
CEDA	Central European Data Agency
CEN	European Committee for Standardization
EIF	European Interoperability Framework
ETL	extract, transform, load
EU	European Union
GDP	gross domestic product
GI	geographic information
GIS	Geographic Information System
GNNS	Global Navigation Satellite System
GUI	graphical user interface
HALE	Humboldt Alignment Editor
IEC	International Electrotechnical Commission
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
OTN	OpenTransportNet
RDBMS	relational database management systems
SDI	spatial data infrastructure
SQL	Structured Query Language
TC	technical committee
TR	technical report
UK	United Kingdom
VGI	voluntary geographic information

Revision History

Revision	Date	Author	Organisation	Description
0.1	12/12/2014	Tomáš Mildorf	UWB	Initial draft
0.2	09/01/2015	Karel Jedlička	UWB	Data models
0.3	16/01/2015	Jan Ježek	UWB	Harmonisation processes
0.4	19/01/2015	Tomáš Mildorf	UWB	Minor corrections
0.5	21/01/2015	Karel Jedlička	UWB	Section 4.2
0.6	23/01/2015	Pavel Hájek	UWB	Section 4.3
0.7	26/01/2015	Tomáš Mildorf	UWB	Reviewers' comments
0.8	29/01/2015	Karel Jedlička	UWB	GTFS added
1.0	2/2/2015	Karel Jedlička	UWB	Reviewers' comments integration Final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Executive Summary

OpenTransportNet (OTN) is a European project co-funded by the Competitiveness and Innovation Framework Programme. The primary aim of OTN is to support the reuse of geographic information (GI) as an important component of public sector information. This will be achieved through cooperation with key stakeholders and development of data hubs which will serve for multiple use. The OTN project focuses on transport data and transport related applications.

OTN is developing collaborative virtual service hubs that aggregate, harmonise and visualise open transport-related data to drive the rapid creation of innovative applications and services. Data that will be used are from various data providers and have different standards and specifications. The heterogeneity of data can be partly overcome by harmonising data according a unified set of specifications. By this approach, the interoperability of data will be achieved.

There are several ways for geographic data harmonisation. There are tools for automation of the processes. However, user interaction for setting the mapping between the original data and the target data model must be defined by a data expert. Once the mapping is done, the process then can be repeated in an automated manner.

The target data model for transport networks, as a core theme for all OTN's applications, is based on the INSPIRE data specifications, particularly on INSPIRE D2.8.I.7 Data Specification on Transport Networks – Technical Guidelines. Other data models for other transport related data are mentioned in this document.

1 Introduction

1.1 OpenTransportNet

OpenTransportNet's (OTN's) primary objective is to create an ambitious network of collaborative virtual service hubs that aggregate, harmonise and visualise open transport-related data to drive the rapid creation of innovative new applications and services. OTN specifically chose transport as a cross border focus-area because it touches upon almost every facet of 21st century living, making it an ideal target for the creation of solutions that can be enhanced by location based services such as GNNS and VGI resources such as OSM.[1]

OTN service delivery hubs will address the following challenges:

1. need to better aggregate and harmonise data to improve accessibility and use,
2. need to link spatial and non-spatial data to extract value and increase accuracy,
3. need to provide innovators with easier APIs and GUIs to stimulate the creation of new services and commercial opportunities.

OTN innovation hubs will overcome the above challenges by using an automated flexible dataset aggregator to integrate and harmonise transport related data. OTN hubs will combine spatial data (GI), dynamic data streams and non-spatial data and derive insights from the data through visualisation tools and pattern detection algorithms. OTN Hubs will go beyond the state-of-the-art by (a) improving the accuracy of data insights by enhancing knowledge with VGI and (b) deploying a sophisticated access control and identity management system that will manage privacy controls.

OTN will be validated in four pilot locations the UK, Belgium, France and the Czech Republic. The OTN solution will use a social enterprise freemium business model. OTN Lite will be an open service that provides access to open data sets and an innovation environment. OTN Premium will be available for a fee that enables users an access to business incubation.

1.2 The Aim and Structure of the Report

A data interoperability is one of the core requirements of the OTN Hub. It is essential that data integrated in the OTN Hub are interoperable, they can be combined and analysed across various themes. Data interoperability is a complex issue involving different aspects of data such as different data formats, reference systems and conceptual schemas. The aim of the report is to provide a methodology for making data interoperable through data harmonisation into common data models. An important aspect is to keep data in line with existing standards for geographic data including INSPIRE, OGC, ISO and CEN.

The report is divided into 5 chapters:

- Chapter 1 is the introduction of the OTN project and the aim and structure of the report.
- Chapter 2 briefly describes data interoperability issues.
- Chapter 3 describes the methodology for data harmonisation including selected tools for data integration and storage.
- Chapter 4 includes the data models for the transport network data theme. Data models for other data themes will be designed on demand during the course of the project. All the data models will be based on existing and standardised data models.
- Chapter 5 is conclusions.

[1] The transport industry directly employs around 10 million people and accounts for about 5% GDP. The quality of transport services has a major impact on quality of everyday citizen lives. The average household spends 13.2% of its budget on transport goods and services Source: http://europa.eu/pol/trans/

2 Data Interoperability

2.1 What is Interoperability

The OTN Hubs will act as spatial data infrastructures (SDIs). SDI, sometimes referred to as spatial information infrastructures, is generally understood as a computerised environment for handling data that relate to a position on or near the surface of the earth (CEN/TR 15449:2011).

There exist many definitions of SDI. The authors use the definition adopted by the INSPIRE directive. INSPIRE defines SDI as “the metadata, spatial data sets and spatial data services; network services and technologies; agreements on sharing, access and use; and coordination and monitoring mechanisms, processes and procedures, established, operated or made available in an interoperable manner.” (European Parliament 2007).

Interoperability is defined by the International Organisation for Standardization (ISO) as "capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." (ISO/IEC 2382-1:1993).

Recent activity of the European Commission brought to an attention a document describing the European Interoperability Framework (EIF) for European public services. The document highlights the needs and benefits of interoperability. Interoperability enables (European Commission 2010):

- cooperation among public administrations with the aim to establish public services;
- exchanging information among public administrations to fulfil legal requirements or political commitments;
- sharing and reusing information among public administrations to increase administrative efficiency and cut red tape for citizens and businesses. (p. 2)

EIF distinguishes four levels of interoperability including legal, organisational, semantic and technical. These levels are described in the deliverable D2.1 Current Situation Analysis (Section 4.2).

2.2 How to Achieve Interoperability

Interoperability on all levels can be achieved through common standards, specifications and other agreements. If all data, services, legislation, technologies etc. share the same set of agreements, the interoperability can be achieved. The most important international and well respected standards in the geospatial domain are from the Technical Committee 211 of the International Organization for Standardization (ISO/TC 211) and Open Geospatial Consortium (OGC). Together with national standards, they create the core for SDI implementation. It is highly recommended to keep the national standards compliant with ISO and OGC standards to enable interoperability across national borders.

ISO/TC 211 Geographic information/Geomatics is responsible for the ISO geographic information series of standards. These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analysing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations. The ISO/TC 211 standards provide a framework for the development of sector-specific applications using geographic data.

The **Open Geospatial Consortium (OGC)** is an international industry consortium of 467 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services and mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications. OGC standards are developed in a unique consensus process supported by the OGC's industry, government and academic members to enable geoprocessing technologies to interoperate, or "plug and play". (Open Geospatial Consortium 2012).

As OTN is dealing also with non-spatial data and Web platforms, the **World Wide Web Consortium** (W3C) standardisation organisation should be mentioned. W3C is developing a set of standards for semantic Web.

The W3C and the OGC announced in January 2015 a new collaboration to improve interoperability and integration of spatial data on the Web.[2]

Successful examples of SDI implementations on international level are described in D2.1 Current Situation Analysis (Section 3.2). Section 3.3 of D2.1 describes standards important for the OTN project.

3 Data Harmonisation and Integration

3.1 Harmonisation Processes

Data harmonisation is necessary for combining data from heterogeneous sources (e.g. regional datasets) into integrated, consistent and unambiguous information products (e.g. European datasets). Such datasets can be then easily used in combination with other harmonised data for viewing as well as querying and analysing. Data harmonisation is a complex task that has not a universal solution that can cover all possible scenarios. Ideal technical solution (system architecture, software) is always determined by many specific facts such as the way how the original data are stored, data volume and the type of harmonisation. The harmonisation process is a best practice in the geoinformation domain and therefore following chapters firstly describes harmonization experiences of the UWB team, acquired during previous projects (Humboldt, Plan4all and Plan4business projects) and formulated in Janečka et al. (2013) into a 5-step harmonisation approach.

All relevant steps to perform data harmonisation are depicted in Figure 1.The first three steps are common steps for all scenarios. The theory of spatial data harmonisation within the framework of INSPIRE is based mainly on the INSPIRE conceptual models. The understanding of both source and target data is based mainly on particular data specifications, documentation and metadata.

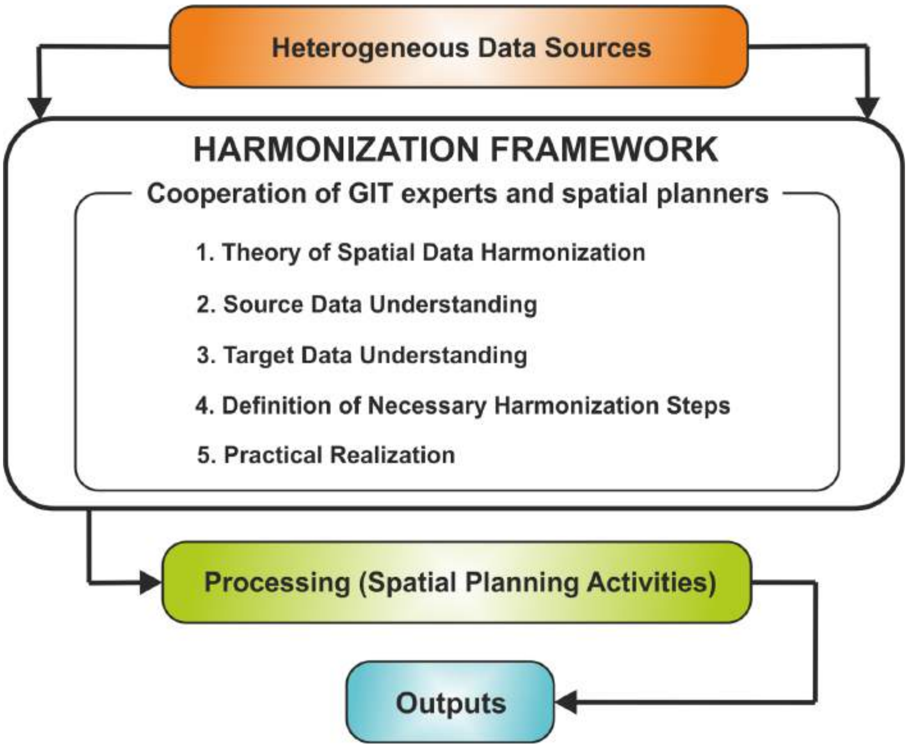

Figure 1. Data harmonisation process (Janečka et al. 2013)

3.2 Harmonisation Tools

From the technical point of view there exist several ways how to handle data harmonisation. Relevant approaches are based on ETL tools (e.g. FME[3]) or specialised software designed for the harmonisation (e.g.

[2] http://www.w3.org/2015/01/spatial.html.en

HALE or Shape Change[4]). Other solution is to use the capabilities provided by relational database management systems (RDBMS) like PostgreSQL with PostGIS. Another solution could be to use a geographic information system such as ArcGIS. In the upcoming sections we mention the tools most relevant for the use within the OTN project. Adapted from Janečka et al. (2013)

3.2.1 HALE

The Humboldt Alignment Editor (HALE) is an open source software framework that was designed in the scope of the Humboldt project. “HALE is a tool for defining and evaluating conceptual schema mappings. The goal of HALE is to allow domain experts to ensure logically and semantically consistent mappings and consequently transformed geodata. Furthermore, a major focus is put on documentation of the schema transformation process and its impacts, e.g. in the form of lineage information attached to the resultant transformed data.”[5]

For advanced harmonisation projects, where a collaboration over a large community is required, there is a platform where professionals develop and share data transformation processes. The platform is available at www.wetransform.to.

3.2.2 Relational Database Tools

Harmonisation frameworks focus on setting up the harmonisation rules and maintaining the harmonisation lifecycle. Data harmonisation can be also performed by using a database technique and build in functions.

When we focus on the sub-process that deals with conceptual schema transformation, we can find that RDBMSs can offer capabilities to deal with this issue. Once we are able to import our data to RDBMS, we can utilise existing functions and SQL language to fulfil harmonisation processes focused on change of taxonomies and any related harmonisation of attributes (rename, retype) as well as geometry processing (depending on available spatial functions of particular RDBMS).

3.2.3 GIS Tools

The primary objective of GIS tools is not on data harmonisation. However, GIS tools can be used for this purpose. On the one hand, GIS tools are more suitable for harmonisation of geometries than the relational database tools. On the other hand, typical database tasks such as attribute mapping are more difficult in a GIS tool than in a database. GIS tools also follow the geoprocessing principle during the data processing: “input -> operation -> output” and therefore they produce many temporary layers. Therefore, the user has to take care about naming conventions and data management during the process. ArcGIS (with Arc Toolbox, ModelBuidler and Python) can serve as a GIS example of harmonisation tools.

3.2.4 Short Comparison of Harmonisation Tools

All three above mentioned harmonisation tools can handle both spatial and non-spatial data - even if first two are primarily focused on attribute schema mapping and the third one is more focused on geometry. What do all the three tools have in common is that once the data harmonisation schema is set up, then it could be written as a batch and run in an automated way - with little or no knowledge of the inside of the routine. Practical example of data harmonisation in OTN is shown in Chapter 4.

4 Developing and Populating OTN Data Structures

The results of data harmonisation process applied in the OTN Hubs are data structured according to the OTN user needs. These needs were identified and specified in D2.3 - Pilot Scenarios, then further elaborated in D2.4 - Co-Design Workshop Reports and incorporated into the OTN Hub architecture in D2.5 - OTN Project Architecture Blueprint and Hub Technical Specification.

As OTN primary focus in on transport, the initial dataset that was harmonised into the OTN Hub are transport networks. Other data than transport networks will be harmonised during the course of the

[3] http://www.safe.com

[4] http://shapechange.net/

[5] http://community.esdi-humboldt.eu/

project. The workflow for transport networks harmonisation described in this chapter will be a guidance for other harmonisation efforts.

4.1 Transportation Network Harmonisation Process

4.1.1 Understanding Source Data - OSM

The OTN project has four pilot regions including Antwerp, Birmingham, Issy and Liberec Region. According to the second step of the 5-steps harmonisation data approach, which is the understanding of source data, it was necessary to gather all accessible information about the existing transportation related data at pilot sites.

First of all, metadata about the transport related and other datasets were identified and described (see the deliverable D2.1 Current Situation Analysis). Moreover, a survey on existing transportation networks data sets, routing and volumes of traffic in pilot sites' was performed. This was done in the frame of WP4 - Management of Data, Task 4.1 - Initial Pilot Dataset Identification and Task 4.3 - Wider Data Collection and Publishing Strategy. The results are presented in Table 1. More details are described in D4.2 - Data Collection and Sharing Plan - Annex 2, Section 9.1.

		Antwerp B	Birmingham GB	Issy F	Liberec CZ
Road network	data on the notwor	✓ yes	✓ yes	✓ yes	✓ yes
	vector data	✓ yes	✓ yes	✓ yes	✓ yes
	GIS format	✓ yes	✓ yes	✓ yes	✓ yes
	data on road sectio	✓ category	✓ category, direction	✓ category	✓ category, tech. paramete
	open data	✓ yes	• not identified	✓ yes	• not identified
	note		*only Open street map*		*Road databank*
Volume of traffic	data on volumes of	✗ no	✓ yes	✓ yes	✓ yes
	vector data	✗ no	✓ yes	✓ yes	✓ yes
	lino/point	✗ no	✓ points	✓ lines	✓ lines
	GIS format	✗ no	✓ yes	✓ yes	✓ yes
	monitored area	✗ no	✓ city	• not identified	✓ district
	necessary data	✗ no	• not identified	✓ yes	✓ yes
	categories of vehic	✗ no	✓ cars, trucks, cyclists	• not identified	✓ cars, trucks, cyclists etc.
	surveys	✗ no	✓ detection	• not identified	✓ manual/detection
	open data	✗ no	✓ yes	✓ yes	✓ yes
	note		*portal WMTIG*		
Demography	demographic data	✓ yes	✓ yes	✓ yes	✓ yes
	territorial division	✓ building blocks	✓ areas of 100 housing ur	• not identified	✓ village, basic administra
	detail data	✓ sufficient	✓ sufficient	✓ sufficient	✓ sufficient
	vector data	✓ yes	✓ yes	✗ no	✓ yes
	GIS format	✓ yes	✓ yes	✗ no	✓ yes
	open data	✓ yes	✓ yes	✓ yes	✓ yes
	note		*Mapinfo*	*format XLS*	
Traffic model	model	✓ yes	✓ yes	✗ no	✗ no
	range (size)	• not identified	✓ city	✗ no	✗ no
	uni/multimodal	✓ multimodal	✓ multimodal	✗ no	✗ no
	software	• not identified	✓ VISUM	✗ no	✗ no
	note	*Vlaams Verkeerscentrum*			

Table 1. Summary of the questionnaire on traffic volume data

The survey revealed that three of four pilots plan to use OpenStreetMap (OSM) as a source data for transportation network (Birmingham, Issy and Liberec[6]). The city of Antwerp has got a permission to openly use transport network data of the Flemish Region. Therefore, the OSM was selected as the first source for data harmonisation.

[6] The Liberec pilot still considers to use data from CEDA - but still deals with the licensing issues.

The OSM data model (source) was analysed and various data exports (namely geofabrik.de, OSM2PO and raw XML export) were discussed. As the PostgreSQL with PostGIS extension database server has been set-up as a data storage for the OTN Hub, the OSM2PO tool, which is a PostGIS extension, was used for extracting source OSM data. The structure of extracted data is depicted in Figure 2.

eu_2po_4pgr	
id	INTEGER
osm_id	BIGINT
osm_name	CHARACTER VARYING
osm_meta	CHARACTER VARYING
osm_source_id	BIGINT
osm_target_id	BIGINT
clazz	INTEGER
flags	INTEGER
source	INTEGER
target	INTEGER
km	DOUBLE PRECISION
kmh	INTEGER
cost	DOUBLE PRECISION
reverse_cost	DOUBLE PRECISION
x1	DOUBLE PRECISION
y1	DOUBLE PRECISION
x2	DOUBLE PRECISION
y2	DOUBLE PRECISION
geometry	USER-DEFINED

Figure 2. Structure of OpenStreetMap data downloaded from the OSM servers and stored in PostGIS.

4.1.2 Understanding the Target Data Schema and Setting Up the Mapping

As the user requirements from all pilots mentioned routing as a fundamental need, a target data schema respecting linear topology was sought-after. Finally, the INSPIRE Transport Networks data model was chosen as the harmonised data schema, because it addresses the linear topology and is compliant with the EU legislation. The INSPIRE Transportation Networks schema was analysed and then data structures necessary for routing (RoadLink and RoadNode) were selected (see Table 1). Then the mapping function from OSM data to the INSPIRE Transportation Networks schema was built.

RoadLink featureType					
Atributte	**Type**	**Source**	**Cardinality**	**Stereotype**	**OSM Source**
inspireID	Identifier	Network::NetworkElement	[0..1]		OSM.roads.osm_id
beginLifeSpanVersion	DateTime	Network::NetworkElement	[1]	voidable, lifeCycleInfo	
endLifespanVersion	DateTime	Network::NetworkElement	[0..1]		
centerlineGeometry	GM_Curve	Network::Link	[1]		OSM.roads.geometry
fictitious	Boolean = false	Network::Link	[1]		false
direction	sign	LinkDirectionValue	[0..*]		OSM.roads.oneway
validFrom	DateTime	Common Transport Elements::TransportLink	[1]	voidable	
validTo	DateTime	Common Transport Elements::TransportLink	[0..1]	voidable	
geographicalName	GeographicalName	Common Transport Elements::TransportObject	[0..1]	voidable	OSM.roads.name

RoadNode featureType					
Atributte	**Type**	**Source**	**Cardinality**	**Stereotype**	**OSM Source**
inspireID	Identifier	Network::NetworkElement	[0..1]		OSM.roads.osm_id
beginLifeSpanVersion	DateTime	Network::NetworkElement	[1]	voidable, lifeCycleInfo	
endLifespanVersion	DateTime	Network::NetworkElement	[0..1]		
geometry	GM_Point	Network::Node	[1]		OSM.roads.geometry
validFrom	DateTime	Common Transport Elements::TransportNode	[1]	voidable	
validTo	DateTime	Common Transport Elements::TransportNode	[0..1]	voidable	
geographicalName	GeographicalName	Common Transport Elements::TransportObject	[0..1]	voidable	
formOfRoadNode	FormOfRoadNodeValue	RoadNode	[1]	voidable	

Table 2. RoadLink and RoadNode structure

The integral part of defining the target schema is mapping it to the source data. As can be seen from Table 2, all mandatory target fields can be populated from the source data. Some of them have to be even deeper specified:

- **RoadLink.centerlineGeometry** needs topologically clean geometry.
- **RoadLink.direction** is filled by **LinkDirectionValue codelist**:

Value	OSM source (OSM.roads.oneway)
bothDirections	0
inDirection	1 (follows the way of vectorization)
inOppositeDirection	1 (opposite)

- **RoadNode.geometry** is generated from RoadLink intersections.

Also other attribute fields are useful for user scenarios and therefore there were added to RoadLink featureType even they are properties of TransportProperty class in the INSPIRE Transportation Network schema:

RoadLink featureType					
functionalRoadClass	FunctionalRoadClassValue	Road Transport Network::FunctionalRoadClass	[1]	voidable	OSM.roads.type
formOfWay	FormOfWayValue	Road Transport Network::FormOfWay	[1]	voidable	OSM.roads.type

- **FuncitonalRoadClass** was populated by data according the FunctionalRoadClassValue enumeration:

Value	OSM source (OSM.roads.type)
mainRoad	motorway, motorway_link, trunk, trunk_link
firstClass	primary, primary_link
secondClass	secondary, secondary_link
thirdClass	tertiary, tertiary_link
fourthClass	residential, living_street, unclassified
fifthClass	<all other values>

- FormOfWay was populated by data according the FormOfWayValue codeList:

Value	OSM source (OSM.roads.type)
bicycleRoad	cycleway
dualCarriageway	motorway_link, trunk, trunk_link, primary_link, secondary_link, tertialy_link
enclosedTrafficArea	raceway
entranceOrExitCarPark	<not a corresponding value>
entranceOrExitService	<not a corresponding value>
freeway	<not a corresponding value>
motorway	motorway
pedestrianZone	<not a corresponding value>
roundabout	<not a corresponding value>
serviceRoad	<not a corresponding value>
singleCarriageway	<all other values>
slipRoad	<not a corresponding value>
tractor	<not a corresponding value>
trafficSquare	<not a corresponding value>
walkway	pedestrian, footway, steps, path

4.1.3 Performing the Transformation Process

The data harmonisation process was done in two steps:

1. Already mentioned import of routing data from OSM to PostGIS database was done by the OSM2PO tool in the phase of understanding of the source data (Section 4.1.1).
2. Data were converted into the INSPIRE-based database schema (described in 4.1.2) using PL/pgSQL functions. The result is the physical schema of spatial database stored in PostGIS and depicted in Figure 3.

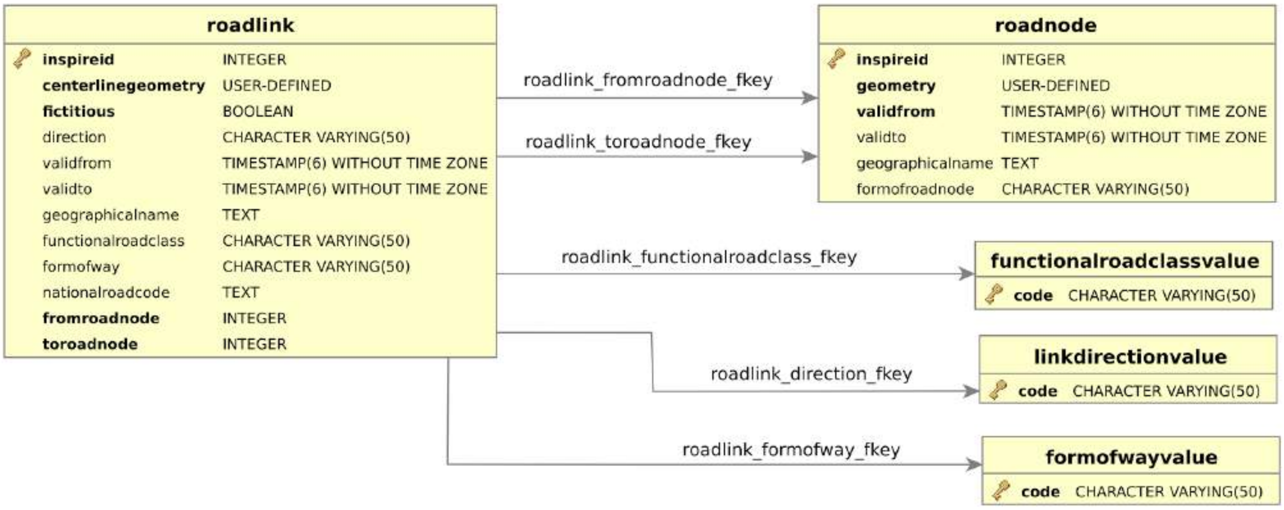

Figure 3. The diagram of tables stored in database.

Finally, there was a script in the PLPGSQL procedural language implemented that performed the batch-processing of all the available data.

In the scope of this harmonisation, the OSM road network that covers the whole Europe was imported to the database. As the next step, Areas of interests of each particular pilot sites can be then extracted if necessary. If a pilot decides to use other source data for the transport network theme, the source will be investigated and the mapping process will be adapted to these data by data experts. It is important that the result of the process will be the same (harmonised) data structure as described in Section 4.1.2.

4.2 Traffic Volumes Data Harmonisation and Calculation

A traffic network stored in the routable data structure is the result of the harmonisation process described in Chapter 3. But it can be also a source data for further data processing. It has been agreed in the OTN consortium, that there is going to be created an application which will be common to all pilot cities. This application will portray the dynamics of volumes of traffic during time. Good to know the traffic volumes is particularly true in densely populated areas with big traffic. The vision and principle of the data harmonisation necessary for traffic volumes visualisation is described in the further text, but it was firstly introduced in a paper presented by the OTN team at the ISAF conference in Jelgava, see Kozhukh et al. (2014) for the report.

Traffic volume is a parameter of a road network which describes the amount of vehicles which go through a network segment in a time period. Together with an information about the maximum capacity of network segments, it can be forecasted where the volume of traffic is going to cause traffic disruptions and traffic jams. We can distinguish three types of traffic volumes:

- daily traffic volume (different for each day from Monday to Sunday),
- annual average of daily traffic volume (AADT),
- peak hour traffic volume – in the busiest hour of the day.

Also a long term predictions can be made (using a mathematical traffic model) calculating the traffic volumes 10, 20 or even 30 years into the future.

4.2.1 Understanding Source Data

In general, there are three basic types of data necessary for traffic volume calculation using a mathematical traffic model, according to e.g. Kozhukh et al. (2014):

- traffic generators - demographic data about places that are usually represented as points. These points can be cities, city districts or building blocks – it depends on the granularity of the data and the desired level of detail. These data are used for estimation of traffic flows in the network. Distinguishing between different types of places such as living, industrial, service or shopping place is useful for estimation of traffic flows direction changes in time. Such a typical day distribution of Traffic Volume variation is depicted in Figure 4.

- Road network - well defined and topologically correct road network is the fundamental constraining graph structure, which describes the allowed movements between different places.
- Calibration measurements - physical measurements of traffic volumes (traffic censes) at particular spots of the traffic network are used for calibration of calculated volumes.

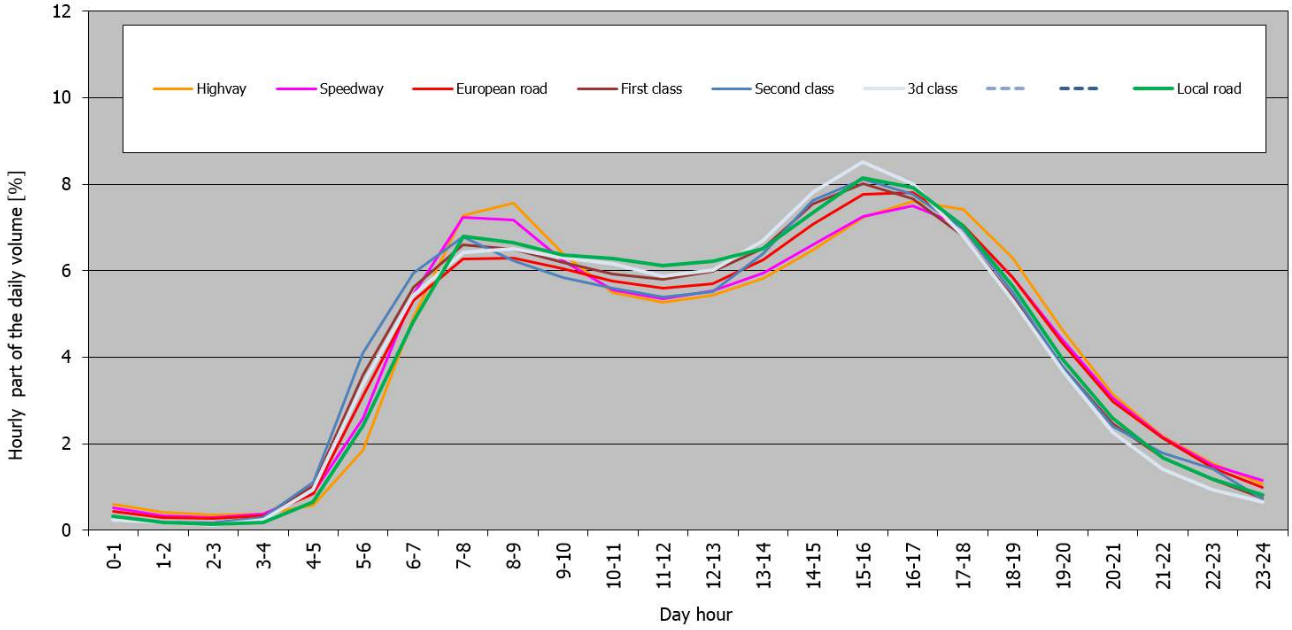

Figure 4. Graph of hourly variation of traffic volume depending on road type (EDIP 2012).

4.2.2 Performing the process of traffic volume calculation

There exist several tools for traffic modelling, for example EMME, CUBE, PTV VISUM, SATURN, TRANSCAD or OmniTrans. All of them are based on similar principles (Kozhukh et al. (2014)):

- First of all, the road network topology and consistency have to be checked (deleting pseudo-nodes, cleaning gaps and overlaps). Then junctions are computed and turns defined. The network is equipped attributes (speed, capacity).
- Then, as the places do not have to lie exactly on a network segment, a connector from each place to the nearest network part (junction or segment) is created. The defined crossing with the network represents a point, in which the people enter the network and generate the traffic.

The two above described points are usually realised in a geographic information system (GIS). The following steps are calculated in a transport engineering software:

- Using the demographic data about traffic generators, various types of traffic volumes are calculated, see for example the work of McShane at el (1990) for more details.
- Then the OD matrix (origin-destination matrix) is calculated and this matrix is “put” on the network. The volume of traffic on each segment of network is calculated. This step produces theoretical volumes of traffic – it can be visualised which road segment has higher traffic volume then the other.
- Afterwards, those volumes are calibrated on real values from traffic censuses.
- The final step is an export from a transportation software to a GIS, where the data can be visualised or used together with the rest of geographic data.

The result of above mentioned process is an average daily traffic volume per each network segment. To incorporate the variation of the traffic volume during the day, the graph of hourly variations (Figure 4) has to be processed - the hourly percentage contribution of Traffic volume is multiplied by each particular segment (see result in Figure 5).

4.2.3 Understanding the User needs and Target Data Schema

The peak and hourly variations of traffic volumes are useful for crisis management (Liberec Region), ordinary routing (Issy-les-Moulineaux), road safety analysis (Birmingham) and redirecting of traffic flows (Anwerpen). Furthermore, city network reconstructions as well as urban planning can take advantage of longer term traffic volume predictions, see more in Martolos & Šindlerová (2013). Therefore, various types of traffic volumes can be calculated in the OTN project as a unifying theme which naturally interconnects all four pilots. The enrichment of transportation network data structure by hourly variation of traffic volume is depicted in figure 5. There have been added attributes storing the hourly contribution to daily traffic volume - for every hour in a day. Their values are calculated from the table TrafficVolumesHourlyVariation, which is tabular expression of the graph in Figure 4. Moreover, also an attribute capacity, describing maximum hourly amount of vehicles crossing the RoadLink, has been added.

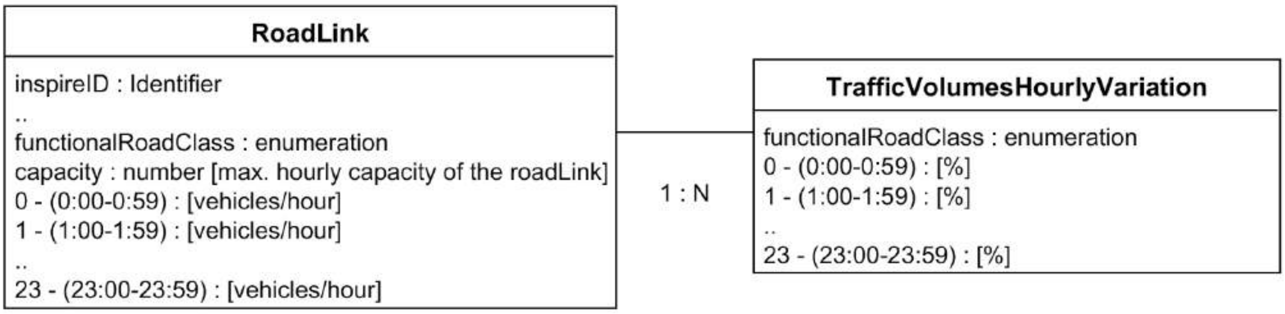

Figure 5. Enrichment of transportation network data structure by hourly variation of traffic volume.

Such a data structure then allows application developers to create an application which dynamically visualize the traffic volume changes. There was already created such an application during the Open Data Hackathon[7] in Jelgava (September 2014) which serves as a proof of concept. See Figure 6 for a screenshot or visit the following URL: http://gis.zcu.cz/projekty/OTN/TrafficVolumesExample.html for a live example. The width of the RoadLink shows the amount of vehicles crossing the segment per hour. The colour shows, how close is the TrafficVolume to the maximum RoadLink capacity (green = 0 % - 50 %, yellow = 50 - 70 %, red = more than 70 %. Note also the time slider, which allows the user to see the data in various times.

[7] http://odh.isaf2014.info/projects/open-data-hackathon/wiki/Open_Data_Hackathon_Wiki

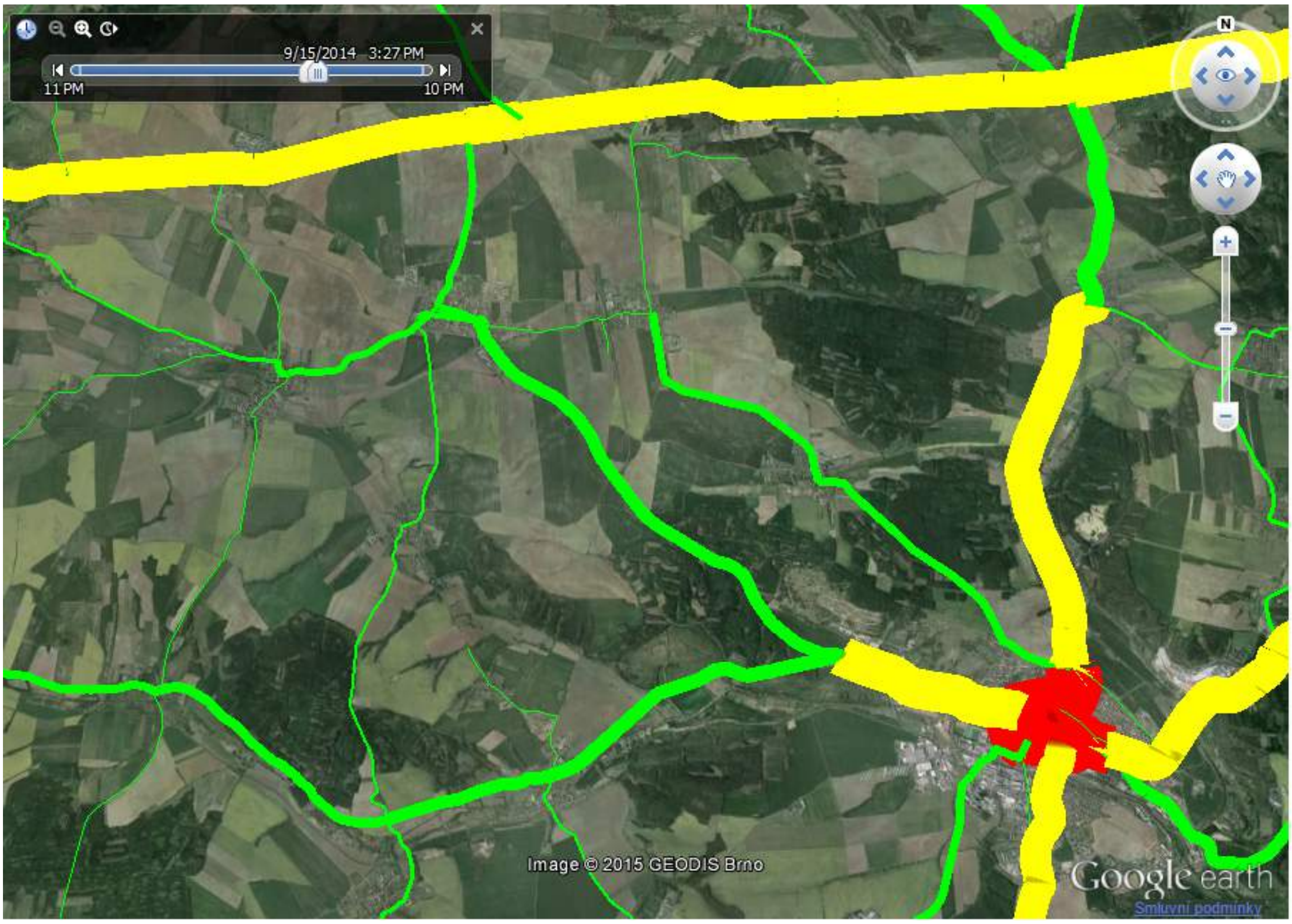

Figure 6. Screenshot of the *Day time variation of traffic volume* application.

4.3 DATEXII

As the DATEXII is the source format of data from Birmingham pilot, its short description follows (adopted from the DATEXII official web site[8]):

Delivering European Transport Policy in line with the ITS Action Plan of the European Commission requires coordination of traffic management and development of seamless pan European services. With the aim to support sustainable mobility in Europe, the European Commission has been supporting the development of information exchange mainly between the actors of the road traffic management domain for a number of years. In the road sector, the DATEX standard was developed for information exchange between traffic management centres, traffic information centres and service providers and constitutes the reference for applications that have been developed in the last 10 years. The second generation DATEX II specification now also pushes the door wide open for all actors in the traffic and travel information sector.

DATEX II is a multi-part Standard, maintained by CEN Technical Committee 278, CEN/TC278, (Road Transport and Traffic Telematics), see www.itsstandards.eu. The first three Parts of the CEN DATEX II series (CEN 16157) have already been approved as Technical Specifications. These three Parts deal with the most mature and widely used parts of DATEX II: the modelling methodology (called Context and framework) as Part 1, Location referencing as Part 2 and the most widely used DATEX publication for traffic information messages (called Situation publication) as Part 3. For the meaning of the term DATEX see Figure 7.

[8] http://www.datex2.eu/content/datex-background

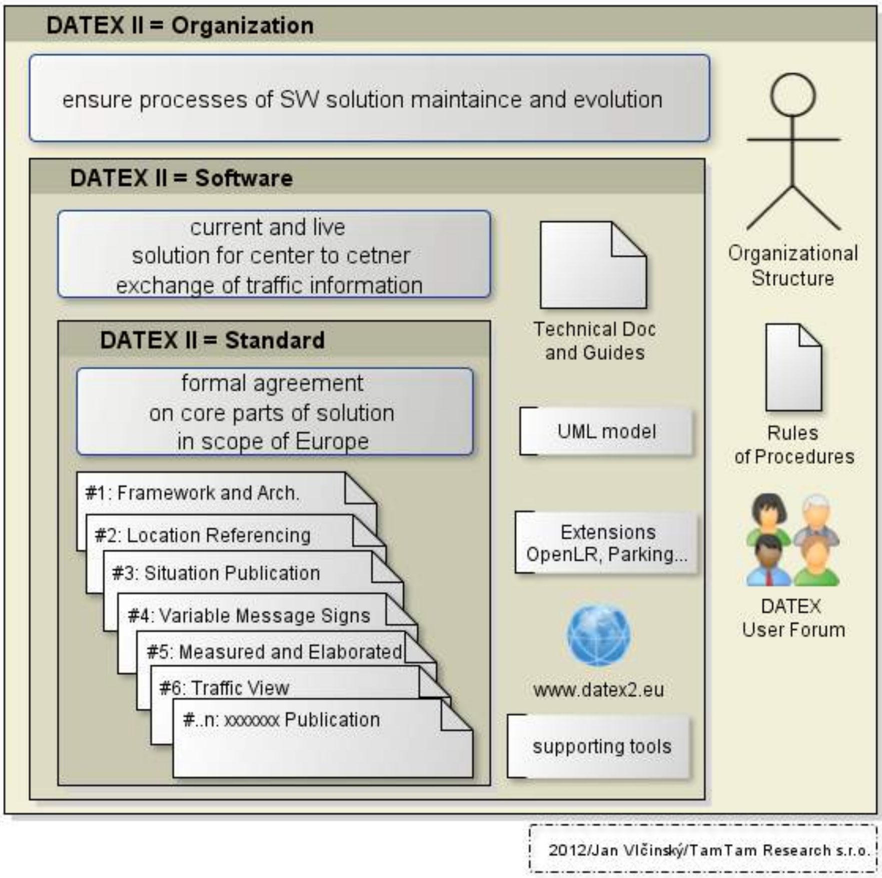

Figure 7. Figure with the schema of DATEX format (http://www.datex2.eu/).

4.3.1 Understanding Source Data of Birmingham Pilot

The data obtained from the pilot area of Birmingham is provided in the XML format in the DATEXII form. The provided data show traffic flows recorded by camera sensors. The measurements of sensors are stored within *measurementSiteRecord* tags and the list of sensors is stored within *MeasurementSiteTable* tags. But not all sites have the sensors installed.

Even though the XML is very versatile language, it would be better to have data in a geospatial format for easier understanding and reusing of data. In this way, it would be necessary to parse the XML, create the relations between the mentioned tables and the spatial reference of the sensors to harvest the structure of data useable for the purposes of the OTN project. While having the data in a geospatial data format, the spatial operations on data would be easier.

4.3.2 Understanding the Target Data Schema

The target data schema for the data stored in DATEX format will be similar to the data schema mentioned in Section 4.2.3. The different approach for the road safety analysis of Birmingham is in using the real measured traffic flow (traffic volumes respectively) instead of the calculated values of the traffic volumes according the road network and traffic generators (as depicted in Section 4.2.1).

Another step of using the provided data is in creation of relation of this data with another data set of roads segments (i.e. the localisation of the sensors within the road network), because the spatial location of the sensors should be in the vertices of the linear roads segments.

4.4 Transmodel for Public Transport Information

Transmodel is a European standard EN 12986 describing the European reference data model for public transport information. The Transmodel standard provides a framework for defining and agreeing data models, and covers the whole area of public transport operations. The Transmodel standard is predominantly used by organisations within the public transport industry that specify, acquire and operate information systems and organisation that design, develop and supply information systems for the public transport industry.

In order to achieve data interoperability in the OTN Hub, this standard will need to be considered for data integration into the data storage.

The data model for network description is a reference data model including entity definitions for different types of points and links as the building elements of the topological network. Stop points, timing points and route points reflect the different roles one point may have in the network definition, for example as a location against which timing information such as departure, passing, or wait times are stored in order to construct the timetables.

The line network is the fundamental infrastructure for the service offer that is provided in a form of vehicle journeys which passengers may use for their trips.

The functional views of the network are described as layers. A projection is a mechanism enabling the description of the correspondence between the different layers. This mapping between the layers is particularly useful when spatial data from different environments have to be combined. An example of such a situation is the mapping of the public transport network on the road network.

The Geographical Data Files (GDF) standard includes a data model for the geographical description of road networks. The GDF standard is defined by the EN ISO 14825:2011 standard. It provides a basic network description upon which various layers describing specific aspects of the use of the infrastructure network may be placed. There is a defined interface for the purpose of data exchange between a GIS and a public transport application.

4.5 General Transit Feed Specification

Transmodel, even it is European standard, is not the only way to deal with data related to public transformation. The General Transit Feed Specification (GTFS), format specified by Google, should be also taken into account.

GTFS defines a common format for public transportation schedules and associated geographic information. GTFS "feeds" allow public transit agencies to publish their transit data and developers to write applications that consume that data in an interoperable way. A GTFS feed is composed of a series of text files collected in a ZIP file. Each file models a particular aspect of transit information: stops, routes, trips, and other schedule data. [9] The basic structure of GTFS data is described at https://developers.google.com/transit/gtfs/examples/gtfs-feed. The complete reference of the format is available at https://developers.google.com/transit/gtfs/reference. A wide range of companies using the format can be explored at https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds.

4.5.1 Comparison with Transmodel

On the one hand, the advantage of GTFS in comparison with Transmodel is that GTFS gives free access to the format reference (see the links above). On the other hand, the disadvantage of GTFS is missing structures for timing and operational data, Knowles et al. (2009). Knowles et al. (2009) compared the GTFS and the Transmodel conceptual models and from this they proposed a convergence path so that TransXChange and NeTEx would be fully compatible and interoperable with GTFS. Their comparison was built upon the GTFS / Transmodel Comparison & Schema published by Kizoom and Miller (2008).

4.6 Other Data Models

Other data models related to transport are included in the deliverable D4.2. Data Collection and Sharing Plan (Chapter 4).

[9] https://developers.google.com/transit/gtfs/

5 Conclusions

The deliverable includes an approach for data harmonisation and storage which should be followed when integrating data into the OTN Hub. The process of data harmonisation and selected tools for harmonising GI data are described. These tools and techniques were developed and thoroughly tested in previous EU projects including Humboldt, Plan4all and Plan4business.

As a result of the discussion of data modelling experts in OTN, a data model for transport network was designed based on the INSPIRE Data Specification on Transport Networks[10]. The OTN transport data model is suitable for routing applications.

Data models for other data are still unclear and will be defined and agreed in a later stage of the project. For this purpose, several transport related data specifications are mentioned.

References

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). Available at: http://eurlex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML [Accessed January 14, 2015].

EDIP, TP 189 Stanovení intenzit automobilové dopravy, [Determination of automobile volume of traffic], 2012, ISBN: 978-80-87394-06-9

International Organization for Standardization, 1993. ISO/IEC 2382-1 Information technology -- Vocabulary -- Part 1: Fundamental terms.

International Organization for Standardization, 2011. CEN/TR 15449 Geographic information - Standards, specifications, technical reports and guidelines, required to implement Spatial Data Infrastructures.

International Organization for Standardization, 2011. ISO 14825 Intelligent transport systems -- Geographic Data Files (GDF) -- GDF5.Janecka, K., Cerba, O., Jedlicka, K., Jezek, J. 2013. TOWARDS INTEROPERABILITY OF SPATIAL PLANNING DATA 5-STEPS HARMONIZATION FRAMEWORK. In SGEM2013 Conference Proceedings, ISBN 978-954-91818-9-0 / ISSN 1314-2704, June 16-22, 2013, Vol. 1, pp 1005 - 1016.

Kizoom, N., Miller, P.. A Transmodel based XML schema for the Google Transit Feed Specification With a GTFS / Transmodel comparison. 2008. http://www.transmodel.org.uk/schema/doc/GoogleTransit/TransmodelForGoogle-09.pdf

Knowles, N., Miller, P., and Drummond, P. Briefing Paper for the Public Transport Coordination Group (PTIC). Version 4 (Issued), 6 February 2009, http://www.rtig.org.uk/web/Portals/0/090206%20-%20PTIC%20-%20Transmodel%20GTFS%20-%20v4%20(Issued).pdf

Kozhukh, D., Jedlička, K., Mildorf, M.,Charvát, K.,Charvát K., Jr., Martolos, J., Šťastný, J., Traffic Volumes Described on Examples in the Open Transport Net Project Pilot Regions . Proceedings of the 18th International Conference on Information Systems for Agriculture and Forestry (ISAF) 2014, Jelgava, Latvia, ISBN 978-80-905151-2-3

Martolos, J., Šindlerová, V. Nová metodika stanovení intenzity generované dopravy. Dopravní Inženýrství. 01/2013. ISSN 1801-8890. Available: http://www.dopravniinzenyrstvi.cz/clanky/nova-metodika-stanoveni-intenzity-generovane-dopravy/ [Accessed: 15 Oct. 2014]

Open Geospatial Consortium, 2012. Open Geospatial Consortium. Available at: http://www.opengeospatial.org [Accessed January 14, 2015].

[10] http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_TN_v3.0.pdf