A vision on transport data sharing in the Smart Cities of Madrid and Ghent

White paper 1.1

Authors:

David Chaves, Pieter Colpaert, Dries Van Ransbeeck, Thimo Thoeye, Ruben Taelman, Peter Bogaert, Javier Chamorro, Oscar Corcho

2015-EU-IA-0037

This document is an official output from the OASIS team to the European Commission. The text will be disseminated as separate vision texts on the website of OASIS, on the blog of open Summer of code, in a paper for the ITS-EU 2017 congress, and on the blog of the Open Belgium 2017 conference.

Contents

Contents **2**

Executive summary **3**

Introduction **5**

State of the art in Transport Data **5**

HTTP as a uniform interface 5

Data publishing formats 5

Semantic interoperability 6

Querying data 8

Exchanging transport data 8

GTFS 9

GTFS-Realtime 10

Linked Connections 10

Transmodel and NeTEx 10

Datex II 11

INSPIRE 11

W3C Semantic Sensor Network 13

Metadata 14

The next steps **16**

Metadata 16

Open Transport Data publishing 17

Developing open systems openly at “open Summer of code” 18

Executive summary

Travelers need route planning advice that crosses the border of single transport modes and operators. We claim that the adoption of an open data policy for transport-related datasets can be highly beneficial for the emergence of all sorts of solutions in this space. In order to stimulate data (re)use, a cost-benefit analysis should be made: if the benefits to reuse these datasets outweigh the costs of making them available, these datasets are going to be picked up at a fast pace. Furthermore, we also want to keep the costs of publishing the data itself at a reasonable level.

We summarize our principles on the sharing of transport data in order to foster (re)use as follows:

1. Datasets must have an **open license**, complying to the Open Definition.
2. **URIs** must be used to identify resources
3. In order to make the datasets more **discoverable**, metadata (using DCAT-AP or possible extensions) must be used to describe the datasets.
4. In order to make data **queryable**, multidimensional interfaces, following the REST constraints for maximum scalability should be adopted.

During the OASIS project, we will:

1. Study the metadata of existing open data portals that publish transport-related data and increase discoverability through DCAT-AP
2. Research on queryable and cost-efficient data publishing:
    a. Work with the Hydra working group to advance the hypermedia specification
    b. Advance the state of the art on multidimensional interfaces within the Linked Data Fragments framework
3. Publish data from
    a. The Flemish bus company De Lijn and the Belgian railway company as Linked Connections (when the data becomes openly available)
    b. CRTM as Linked Connections
    c. CRTM on transport card validation using the W3C SSN Ontology
4. Develop analytical tools for transport networks on top of Linked Connections at open Summer of code
    a. An analytical tool for public transport networks
    b. Estimation of occupancy on vehicles in Madrid

Introduction

Travelers want their route planning advice to include different transport modes and operators. Trying to answer the question “how long do I have to walk from one point to another?” can take into account the geolocation of the streets, the weather conditions at that time of the day, the steepness of the road, whether or not there is a sidewalk, criminality reports to check whether it is safe to walk through these streets, how busy a train or bus is, the accessibility of the road for e.g., wheelchair users, whether the street is blocked by works at that time, etc. We can imagine the complexities that arise when the user does not only want to walk, but that he also wants to get advice when taking into account different private and public transport modes. A data source should thus be able to be handled with the assumption that there is more data outside of this pool that may be relevant to the question. Publishing data for maximum reuse – with this open world assumption in mind – is a challenge of raising interoperability with other datasets, requiring a high availability, a high user-perceived performance, and a good discoverability of the data source.

State of the art in Transport Data

How are transport datasets published today? We will discuss this from two perspectives: data formats as well as the meaning (semantics) of the terms used within these formats. We will discuss the current de-facto standards within the transport domain as the last section.

HTTP as a uniform interface

Data publishing formats

In order to put data in a document, database or datadump, or when transmitting a dataset, a data format or serialization needs to be agreed upon. Examples of common serialization are Comma-Separated Values (CSV), the hierarchical Javascript Object Notation (JSON) or the Extensible Markup Language (XML).

```
"id","starts from"
"Train P8008","Schaerbeek Station"
```

Figure 1: An example dataset describing a train leaving in CSV

As humans decide how datasets are shaped, human language is used to express facts within these serializations. The smallest building block to express a fact one can think of, is a triple, containing a subject, a predicate and an object, in which the subject has a certain relation to an object, and this relation is described by the predicate. In Figure 2, we illustrate how the CSV example in Figure 1 would look like in a triple structure.

```
Train P8008 → starts from → Schaerbeek Station
```

Figure 2: An example dataset describing a train leaving in CSV

This triple structure – rather than a tabular or hierarchical data model – helps studying data using its most basic building blocks. It is also the basis of graph-based data models, as each subject, predicate or object can re-occur in another triple. Different dedicated serializations for triples exist, such as Turtle and N-Triples. Also existing serializations such as JSON, CSV or XML can be used to encode triples within a document.

Semantic interoperability

Let us perform a small thought experiment… Imagine three triples that are published under an open license by three different authorities. One machine publishes the triple in Figure 2, while two other publish the triples illustrated in Figure 3 – any serialization can be used – representing the facts that train P8008 starts from Schaerbeek Station.

```
Schaerbeek Station → Located in → Schaerbeek City

Jacques Brel → born in → Schaerbeek
```

Figure 3. Two other triples published by two different sources

When a user agent, a computer program that acts on behalf of an end-user, visits these three machines, it can now answer more questions than each of the machines would be able to do on their own, such as: “What trains leave in the city in which Jacques Brel was born?”. Nonetheless, a problem occurs: how does this user agent know whether “Schaerbeek City” and “Schaerbeek” are actually the same entity?

Instead of using words to identify things, it is more pragmatic to use unambiguous identifiers, such as a number to identify things. This way, every organization can have their context in which entities are described and discussed. E.g., the station of Schaerbeek could be given the identifier 132, while the city of Schaerbeek could be given the identifier 121. Yet for an outsider, it becomes unclear what the meaning is of 121 and 132, as it is unclear where its semantics are documented, if documented at all.

Resources can be anything, including documents, people, physical objects, and abstract concepts. Within Resource Description Framework (RDF), they can be identified using a Uniform Resource Identifier (URI), or represented by a literal value (such as a date or a string of characters).

Linked Data solves this problem by using Web identifiers, or HTTP URIs. It is a method to distribute and scale data over large organizations such as the Web. When looking up this identifier – by using the HTTP protocol or using a Web browser –, a definition must be returned, including links towards potential other interesting resources. The triple format to be used in combination with URIs is standardized within RDF. In Figure 4, we exemplified how these three triples would look like in RDF.

```
<http://example.org/trains#P8008>
<http://example.org/terms#startsFrom>
<http://irail.be/stations/NMBS/008811007> .

<http://irail.be/stations/NMBS/008811007>
<http://dbpedia.org/ontology/location>
<http://dbpedia.org/resource/Schaerbeek> .

<http://www.wikidata.org/entity/Q1666>
<http://www.wikidata.org/entity/P19>
<http://dbpedia.org/resource/Schaerbeek> .
```

Figure 4. Three RDF triples in the N-Triples serialization using URIs for each term.

The URIs used for these triples already existed in other data sources, and we thus favoured using the same identifiers. It is up to a data publisher to make a choice on which data sources can provide the identifiers for a certain type of entities. In this example, we found WikiData to be a good source to define the city of Schaerbeek and to define Jacques Brel. We however prefer iRail as a source for the stations in Belgium. As we currently did not find any existing identifiers for the train route P8008, we created our own local identifier, and used the domain name example.org as a base for extending the knowledge on the Web.

Querying data

Just like Linked Data insists on using HTTP identifiers, RESTs uniform interface constraint requires that every individual information resource on the Web is accessed through a single identifier – a URI – regardless of the concrete format it is represented in. Through a process called content negotiation, a client and a server agree on the best representation. For example, when a resource “station of Schaerbeek” is identified by the URI http://irail.be/stations/NMBS/008811007 and a Web browser sends an HTTP request with this URI, the server typically sends an HTML representation of this resource. In contrast, an automated route planning user agent will usually ask and receive a JSON representation of the same resource using the same URI. This makes the identifier http://irail.be/stations/NMBS/008811007 interoperable over different clients, as clients consuming different formats can still refer to the same identifier. This identifier is also sustainable, because new representation formats can be supported in the future without a change of protocol or identifier.

In order to navigate from one representation to another, controls are given within each representation. The principle is called hypermedia: once a user agent receives a start URL, it is be able to answer its end-user’s questions by using the controls provided each step of the way.

Exchanging transport data

Organizations usually save their data in conventional databases, CSV formats or plain text files. Generally, these approaches have developed their own data structures and formats to represent the data. If the organizations follow an open data policy it’s normal that the main ways to get access to this data are through APIs provided by organizations or through getting a copy of the files uploaded on their web pages or on their open data portals.

These approaches can result in a lot of problems if we want to develop robust and high quality applications in the public transport domain. If we have a local copy of the schedules from transport organization, we may not show the correct information because we do not necessarily always have an up-to-date version of the files. On the other hand, with the API approach we could solve the outdatedness problem. But if we want to get data from different organizations, we'll have to use different models to represent the data, one per each organization.

Today, transport data is not yet published according to the REST constraints. We outline different specifications which use HTTP mostly as a transport layer only, rather than as the uniform interface. We outline the de-facto standards in Europe: GTFS, GTFS-RT, Transmodel, Datex and INSPIRE. Work has been put already in advancing this state of the art with Linked Connections, Linked GTFS, giving URIs to INSPIRE terms and Linked Datex2. Furthermore, sensors and their associated datacan be described using the W3C SSN ontology.

GTFS

The General Transit Feed Specification (GTFS) is a framework for exchanging data from a public transit agency to third parties. GTFS, at the time of writing, is the de-facto standard for describing and exchanging transit schedules. It describes the headers of several CSV files combined in a ZIP-file. Through a calendar.txt file, you are able to specify on which days of the week a certain service is going to take place during a part of the year. In a calendar_dates.txt file, you are able to specify exceptions on these calendar.txt rules for, for example, indicating holidays or extra service days. Using these two files, periodic schedules can be described. When an aperiodic schedule needs to be described, mostly only the calendar_dates.txt file is used to indicate when a certain service is running. A gtfs:Service contains all rules on which a gtfs:Trip is taking place, and a gtfs:Trip is a periodic repetition of a trip as defined earlier. Trips in GTFS contain multiple gtfs:StopTimes and/or gtfs:Frequencies. The former – mind the difference with a connection – describes a periodic arrival time and a departure time at one certain gtfs:Stop. The latter describes the frequency at which a gtfs:Trip passes by a certain stop. GTFS also describes the geographic shape of trips, fare zones, accessibility, information about an agency, and so forth. We have given URIs to the terms in the GTFS specification through the Linked GTFS (base URI: http://vocab.gtfs.org/terms#) vocabulary. The definitions of the above terms can be looked up by replacing gtfs: by the base URI.

Up to date, many route planners exist in software-as-a-service platforms such as Navitia.io, in end-user applications such as CityMapper, Ally or Google Maps, or as open-source, such as Open Trip Planner or Bliksemlabs RRRR. Other common practices include that an agency, such as the Dutch railway company, the SNCB or the Deutsche Bahn, expose a route planner over HTTP themselves.

GTFS-Realtime

GTFS-Realtime is an extension of GTFS that was developed to add realtime to the model. This specification allows transport organizations to add their real-time updates like: Trip updates (delays, cancellations, changed routes), Service alerts (Stop moved; unforeseen events affecting a station, route or the entire network) or Vehicle positions (Information about transit vehicles including location and congestion level). The way to generate this update is based on Protocol Buffers.

GTFS-realtime is also easy to develop and useful for transport organizations, but maintains similar problems to the basic approach.

Linked Connections

Linked Connections (LC) is a data publishing platform that follows the REST constraints. It is a route planning system, where user agents are expected to execute the route planning algorithm instead of the server. The server then focuses on publishing the data in a favourable manner for user agents that want to plan routes.

Linked Connections includes both a vocabulary to describe connection objects, as well as a server interface specification to describe a time schedule as a list of connection objects. Introduced in 2016, Linked Connections is still at an early stage. More information at http://linkedconnections.org and an opinion piece about this kind of publishing was published at https://pietercolpaert.be/opentransport/2017/01/15/routeplanning-by-government.html.

Transmodel and NeTEx

Public Transport Reference Data Model (Transmodel) is the European Reference Data Model for Public Transport Information developed by the European Committee of Standardization (CEN). Transmodel is a standard that facilitates the interoperability between information processing systems of transport operators and agencies. It is based on using matching definitions, structures and meanings for the data.

Network and Timetable Exchange (NeTEx) is an extension of Transmodel. It was developed for exchanging public transport schedules and related data based on XML. The development is managed by CEN too.

These approaches have been developed following UML and XML formats, without any semantic approach. They can be a good starting point for analyzing what the main features in public transport domain they have taken into account are and what interoperability means for them. At this moment, these developments are not sufficient for the solutions we want to share during this project, where semantic interoperability is essential.

Datex II

DATEX II aims to be a standard way for communications and interchanges of traffic information between traffic centres, service providers, traffic operators and media partners. The specification provides a way to exchange data across boundaries, by which it improves the management of the European road network.

DATEX II is relevant for developments where dynamic information on the transport systems and notably the road system is concerned. The approach is focused on multiple areas. For example, it can be used to exchange information between individual vehicles and traffic, management systems, etc. that are relevant areas for our project, but it can also be used to other that are not so relevant to us (rerouting, network management, dynamic speed limits, etc.). So surely, it is an approach that it is not going to be enough to represent all features (e.g. linked data, accessibility or card validations) we want in OASIS.

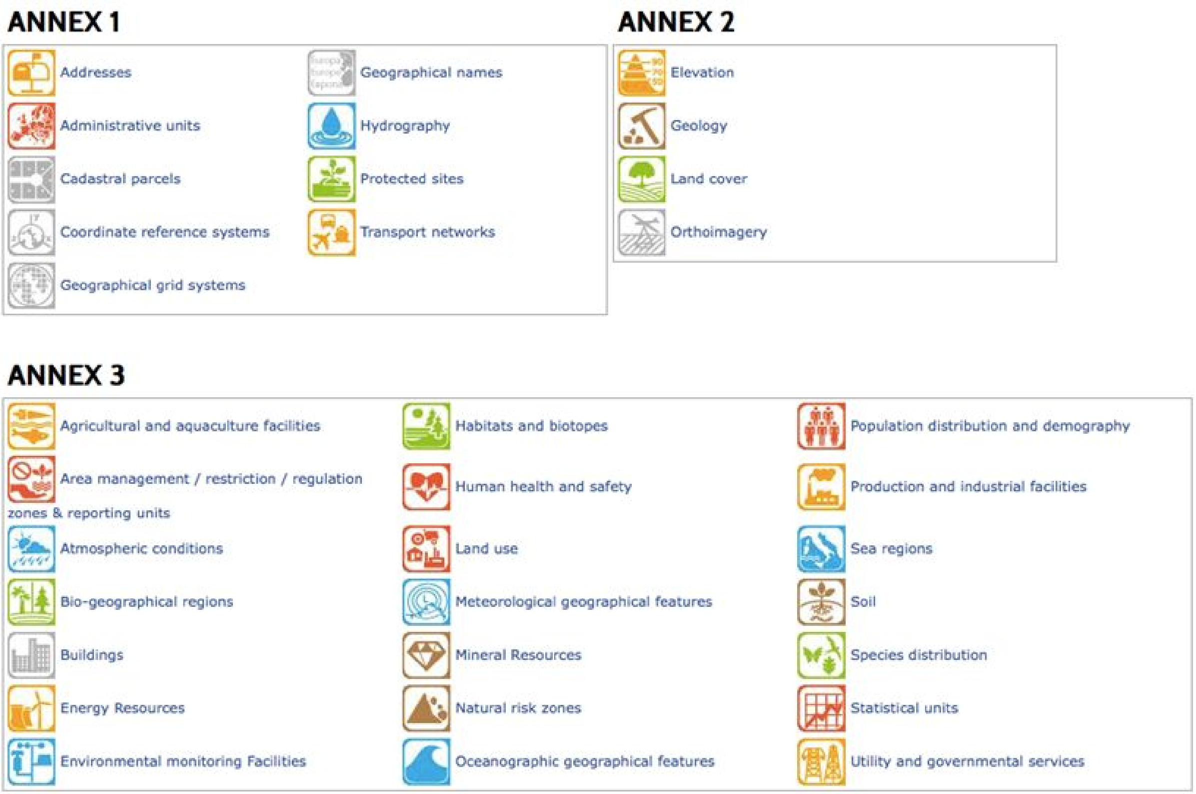
A Linked Data mapping of DATEXII to Linked Datex was created during open Summer of code 2016: http://vocab.datex.org.

INSPIRE

When public governments (or government related bodies) are creating Geospatial data, they have to take into account the specifications defined within the INSPIRE directive.

The main objective of the INSPIRE directive, is making sure that if spatial data exists, concerning one of the 34 themes defined in the directive, it is discoverable and usable in a community and transboundary context. INSPIRE defines the implementation rules in order to set up Spatial Data Infrastructure (SDI) for the European Community. These implementation rules are defined for, amongst others, Metadata, Data Specifications, Network Services, Data and Service Sharing and Spatial Data Services.

Although INSPIRE originated for purposes of EU environmental policies and policies or activities which may have an impact on the environment, the 34 themes in INSPIRE, are found within almost every domain where the location of things plays an important role. Figure 5, gives an overview of different themes defined in INSPIRE.



In relation to Transport, it is obvious that the theme of Transport networks (the location of transport infrastructure and the connection between these locations) is an important one. Also Geographical names (e.g. Station Schaarbeek) Administrative Units (the city of Schaerbeek) are commonly involved.

Thus, when location and governmental data is involved, INSPIRE needs to be taken into account. On the one hand, data managers need clear definitions of what to manage and how it is defined. On the other hand, there is a need to look how the different standards available can be linked to another.

W3C Semantic Sensor Network

The W3C Semantic Sensor Network (SSN) ontology describes sensors and observations, and related concepts. SSN is based on a revised and expanded version of the Sensor, Observation, Sample, and Actuator (SOSA) pattern. SOSA acts as a central building block for the SSN but puts more emphasis on light-weight use and the ability to be used standalone. SSN does not have descriptions about domain concepts, time, locations, etc.

The SSN ontology is designed to offer four identifiable and coherent perspectives or use cases for sensing, each of which may be used either independently or in concert with the others. These perspectives are

- Sensors: with a focus on what senses, how it senses, and what is sensed;
- Data: with a focus on observations and metadata;
- Systems: with a focus on systems of sensors; and
- Features: with a focus on physical features, properties of them, what can sense them, and what observations of them are made.

Currently, the ontology is used in the transport domain to represent the user's card validations at the CRTM services, and it may be useful as well to represent observations related to the positions of vehicles, expected time for arrival of vehicles, etc.

	Schedules	**Realtime**	**Accessibility**	**Traffic**	**Card Validations**	**Geographic Information**
GTFS	X					X
GTFS-Realtime	X	X				X
Linked Connections	X	X				X
NeTEx	X	X			X	X
Datex2				X		X
Inspire						X

W3C SSN					X	

Table 1. Common data formats and their coverage of main transport features we want to represent in OASIS.

Metadata

Metadata are data that describe other data, needed to be able to **discover** datasets within a certain domain. Simplifying, metadata describe the available information about or within a dataset(s). In the transport domain, the metadata files may contain general information like the name of the distribution file or the company web page to whom data belongs. They may also contain more specific information about public transport, like the geographic area covered by a specific service.

One of the most important aspects when open data is published is its discoverability. Taking into account the associated metadata will improve and enrich this task. For example, metadata may allow developers to discover data for their mobility applications, or organizations to know the vocabularies used to describe transport services. In Spain we can find organizations that are related with public transport like CRTM (Transport Consortium of Madrid), EMT (Local Company for Transport in Madrid) or CTZ (Transport Consortium of Zaragoza). These organizations are publishing transport data into their open data portals, but the metadata files do not exist or have been published without following the recommendations of the Spanish Government for metadata production, what makes its discoverability more difficult. We aim to resolve this problem by improving the metadata generation process in public transport organizations by using vocabularies like DCAT or DCAT-AP.

On January 2014, the DCAT W3C recommendation for metadata generation was published. DCAT is a RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. Using DCAT, publishers increase discoverability and applications easily consume metadata from multiple catalogs. Besides, it facilitates federated queries across sites. In Spain, for example, this recommendation is the base for the Interoperability Technical Norm (national recommendation to publish metadata), followed by almost all council portals, like Madrid, Cáceres or Zaragoza.

From a group of European organisations (DG CONNECT, DG DIGIT and Publications Office of the EU), an extension of DCAT has been developed, the DCAT Application Profile (DCAT-AP). This vocabulary includes classes and properties that the initial vocabulary had not taken into account. DCAT-AP is an ontology that contains mandatory, recommended and optional classes and properties, so it's easy to adapt it to different domains. The European Data Portal implements the DCAT-AP as the common vocabulary for harmonising descriptions of over 258,000 datasets harvested from 67 data portals of 34 countries.

The extension is created to find a solution to the problems with the reuse and localization of open data at European level. The DCAT-AP team has identificated two different scenarios, one for data reusers, who find it sometimes difficult to get an overview of which datasets exist and which public administrations are maintaining it and another for data providers, who want to encourage reuse of their datasets by making them searchable and accessible.

In Belgium, a DCAT-AP working group exists at http://dcat.be. It is a working group with all data portal managers from different governments to talk about the adoption of DCAT-AP. The major portals today both have a DCAT-AP harvester, as well publish their raw metadata in this format.

The next steps

We will now provide more details on the work that we will be carry out regarding metadata, open data publishing systems, and how we will create the next proof of concepts at the open Summer of code, for which preparations are ongoing.

Metadata

We defined DCAT and DCAT-AP in the previous section, that are the main vocabularies for making data more visible. The first steps to our metadata assessment will consist of involving some transport organizations to identify what metadata vocabulary is used now, the process of generating that metadata and what the current visibility level of their data is. After that, an analysis will be done where we will try to collect all the common problems associated to the generation of metadata files in the public transport domain.

After our analysis, we will develop a best practices guideline for the generation of metadata in the transport domain. This document will have two clear objectives: the projection of public transport data to European levels and showing public transport organizations the importance of metadata when an open data portal is developed. So, the activities are:

1. Describe the current metadata problems based in the previous analysis and the relevance of taking them into account in the transport domain.

2. Make a version/extension of DCAT-AP focused on the transport domain.

3. Study the feasibility and potential usefulness of developing a specific PublicTransportDCAT Application Profile like others before: GeoDCAT-AP and StatDCAT-AP.

4. Describe the metadata adaptation process followed by CRTM. They will add a DCAT-AP version on top of their current ISO19139 metadata.

Open Transport Data publishing

Open Transport Data publishing entails lowering the cost for reuse by other organizations. One way in which we want to lower this cost is by making our data sources more interoperable. This way, systems will not require new development investments in order for the data to become adopted.

The first requirement for publishing interoperable open data is having an open license. This open license has to be added to the metadata about this resource. We will keep a close eye on whether the metadata also mentions this open license, for *legal interoperability*.

Clients and servers implement the HTTP protocol so that their communication is *technically interoperable*. We will use this HTTP protocol not only as a transfer protocol, but also as the uniform interface to work with the datasets itself. Extra protocols on top of HTTP should focus on the *semantic interoperability* by creating URI vocabularies.

Within OASIS, one of our main goals is to improve the mobility of the citizens from local to European level. To achieve this objective it's important to take into account the *semantic interoperability* to exchange data in an unambiguous way between different transport organizations. With this approach we would solve most of the current problems described before about formats and access. The fundamental idea is to use common formats to represent transport data, as described in the next chapter, and have a standard method to get access to all of these data sources.

Linked Data Fragments (LDF - http://linkeddatafragments.org) is a framework to reuse datasets published with a hypermedia approach. To that extent, a Linked Data vocabulary Hydra is exploited. The Hydra specification is currently a draft at a community group of the World Wide Web consortium and is the vocabulary that powers the querying interface at http://client.linkeddatafragments.org, called Triple Pattern Fragments. We will collaborate with this working group to further develop this specification. More specifically, a simple, yet sufficiently expressive Web API description vocabulary should be finalized so that clients can automatically retrieve data and perform actions that could alter this data. We will do this by analyzing existing Web API's, and seeing which actions they support. This effort will further improve the *technical interoperability* by allowing flexible and automatically derivable communication between clients and servers.

For public transit route planning, the Linked Connections framework will be further developed by integrating it with the Linked Data Fragments ecosystem. Linked Connections requires an ordered array of connection objects, and thus the current LDF client and server need to be extended with multidimensional interfaces. This way, efficient lookups within ordinal ranges become possible, such as geo-spatial search, time window lookups or prefix search, straight from this generic interface.

For other modes than public transport, the biggest problem is that non Linked Data standards are not compatible with Linked Data specifications today, and these non Linked Data standards cannot be used in conjunction with each other. Therefore, we will reach out to the Datex2 community and study whether they can offer URIs for the terms they define. We can then use existing publishing mechanisms over HTTP to make transport data available for maximum reuse.

Developing open systems openly at "open Summer of code"

All data and applications within the scope of the OASIS project follow an open data and open software development philosophy. Therefore, the programming languages, frameworks, libraries and platforms of all applications developed are going to be published under an open license. This allows the reutilization, modifications and exploitation of these developments. As the applications are focused on the transport domain, most of them will be used with smartphones or portable devices (like tablets). The distribution channel will be completely free for final users (e.g. Apple Store, Google Play). The developers and project team will make use of these channels to collect feedback from the users and make updates and improvements.

All used formats have an open nature (e.g. RDF, OWL or LC). These formats will contribute to ease data access through open data portals from public transport organizations that will get into the project (similar to what CRTM is doing now).

All applications linked to this project will be developed during the open Summer of code in Belgium in 2017 and 2018. Within the scope of the action, at least two teams will be recruited during the open Summer of code editions of 2017 and 2018; specifically for the creation of the demonstrators that build on top of the linked open data published by the project, through the generic services developed by the project. The goal of the activity is twofold: on the one hand, this approach will prove that the Linked Open Data published through OASIS is indeed easy for non-expert developers that have no prior knowledge of the project and technologies used therein; on the other hand, it will result in actual production-grade products, that provide a real cross-border and cross-sector service to European citizens.

Thanks to the organisation of many projects and events, among the Apps for Europe project, Open Knowledge Belgium has access to an extensive network of developers and SMEs using open data throughout Europe. During two editions of the open Summer of code (in 2017 and 2018), Open Knowledge Belgium will engage developers to create innovative end-user application for (local) public services and public transport.

Open Summer of code is a summer programme that gives students with development, communication, business and design skills summer jobs with a real challenge: transforming open innovation projects into powerful real-world services for organisations and companies. From building native and web applications to creating new digital services, data visualisation or new datasets. Open Summer of code is intended as a setting in which policy makers, companies and civic technologists intersect, collaborate and start prototyping the future. Inherent to the open Summer of code methodology is the notion that developer teams are not bound to a strict specification of the end-product. Instead, the teams co-create the actual products together with its relevant stakeholders in an agile way.

With the help of the open Summer of code, OASIS will set an example of how cutting edge technologies can be employed in a real-life setting, and how they can indeed improve the mobility of Europeans and the accessibility of European government. The project will deliver, under this activity, at least two demonstrators that work in 2 European countries (Spain and Belgium) using at least public transport data and data about (other) local public services each.

Two specific open-source tools for route planning that are going to be developed during open Summer of code 2017:

- Analytics tool, using analytical queries, to provide insights into transport networks on top of the data from Madrid and Belgium, republished as Linked Connections.

- Tool to estimate the occupancy of public transport vehicles, based on transport card validation data from CRTM, republished using the SSN ontology.

This activity will result in useful and easy-to-use, production-grade, cross-border and cross-sector end user applications (or services) that consume Linked Open Data about public (transport) services. These applications will be completely open source, include extensive documentation and will be ready to be re-used by developers and governments throughout Europe.

By the end of this activity, the consortium aims to have engaged between 8 and 20 developers to work with the Linked Open Data provided and have developed at least 2 applications using the linked open data provided within the open Summer of code.