

A semantic framework for public services and transport data interoperability

Public Services and Transport Ontologies

Authors:

Julian Rojas, David Chaves-Fraga, Pieter Colpaert, 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](#).*

Contents

Contents	2
Executive summary	3
Introduction	4
Public Transport Ontologies	5
Linked GTFS	5
Transport Accessibility extension	5
Linked Connections	6
Mapping Tools	6
Public Services Ontologies	7
Core Public Service Vocabulary Application Profile	7
Infrastructure for Spatial Information in Europe	8
OSLO Building Vocabulary	9
Wheelchair Accessibility Vocabulary	10

Executive summary

The challenges regarding the lack of availability, quality, organization, accessibility and sharing of data are common to a large number of activities and are experienced across the various levels of public authority in Europe. However, thanks to the Open Data initiatives within the European Union the gaps regarding availability of data are being bridged but raising, at the same time, new challenges in terms of interoperability. On this context, interoperability is understood as the possibility to combine data and services from different sources in a consistent way without involving specific efforts of humans or machines. In order to achieve this, common knowledge bases that model the different concepts, elements and properties inherent to the various domains of the public sector are needed.

On this document we present a set of vocabularies aimed at lowering data adoption costs and increasing interoperability by establishing common semantics for the public transport and the public services sector. Such vocabularies were put to test through the development of two different proof of concept application for each domain: Tripscore¹ and AccessFlanders².

The vocabularies that belong to the public transport sector were developed by our team following the model proposed by the currently available defacto standard to share public transport data, due to the lack of an existent vocabulary that covers the main aspects of public transport networks while following the Linked Data principles. As for the public services sector, there are already multiple efforts towards creating common knowledge bases in the European context, which allowed us to reuse and extend these vocabularies according to the particular needs of the sector and the use cases we aimed on.

¹ <http://tripscore.eu/>

² <https://accessflanders.be/>

Introduction

Lowering the adoption costs of open datasets is one of the main challenges for achieving higher interoperability and fostering the creation of new and innovation-driven data ecosystems that allows the development of richer services. The public transport sector can be taken as a clear example of such scenario. Today more and more public transport data are being published as open data aiming for the creation of new and innovative services that address specific needs. However, due to the existence of different data models and formats, creating applications that are able to integrate data from multiple sources demands a high development and investment wise effort. To tackle this issue, the Linked data principles³ propose to use universal identifiers for data through IRIs and to provide common semantics that allow to model and represent specific domain concepts, elements and properties unambiguously. Using standardized formats as RDF and OWL is possible to define reusable vocabularies or ontologies that model specific domains as the public transport sector.

Depending on the use case, semantic ontologies can be extended for more specific goals. For instance, on the public transport sector route planning is a common use case that may involve the integration of multiple datasets (multimodal route planning) and thus require not only to model basic elements like stations and vehicles but also behaviour related concepts like schedules, routes, delays, among others. Another example can be seen on providing information about wheelchair accessibility of public service buildings where physical properties and features that are specific for wheelchairs need to be semantically described among buildings. Taking into accounts these scenarios we developed two proof of concept applications that reuse semantic vocabularies to bring together multiple datasets from different sources and reduce inherent data adoption costs.

On this document we present a description of such semantic vocabularies which were defined by our team according to the use cases needs and also were selected and reused from the Linked Data and semantic Web community.

³ <https://www.w3.org/DesignIssues/LinkedData.html>

Public Transport Ontologies

In this section we present the vocabularies defined during the project to describe and represent public transport networks and that are used in the TripScore application.

Linked GTFS

The General Transit Feed Specification (GTFS)⁴ 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 on the Web. In OASIS, we reused the terms and definitions of GTFS to create the Linked GTFS vocabulary⁵ in order to provide a common description to elements, properties and concepts that belong to the public transport sector. For example:

- **gtfs:StopTime** - Marks the time at which a given vehicle will stop at a certain location.
- **gtfs:Trip** - Represent a collection of gtfs:stopTimes that are followed by a given vehicle.
- **gtfs:bikesAllowed** - Indicates whether bikes are allowed in a certain vehicle or not.

The Linked GTFS vocabulary provides a comprehensive model for describing common concepts of the public transport sector and can be reused to facilitate interoperability among public transport datasets from different sources.

Transport Accessibility Extension

During the Open Summer of Code 2018 carried out in Madrid (Spain) an extension of the Linked GTFS vocabulary was created. This extension aims on describing different accessibility features for public transportation, more specifically on Bus and Metro systems. The extension introduced concepts and definitions like the following:

- **trac:Elevator** - Represents a physical elevator that may be present in stations as a mean of access.
- **trac:DisabilityFeature** - Represents any feature available in the a transport network aimed to address a given type of disability.
- **trac:DisabilityType** - Specifies the category of a disability. E.g. visual, hearing, cognitive, motor, etc.

⁴ <https://developers.google.com/transit/gtfs/reference/>

⁵ <http://vocab.gtfs.org/terms#>

This vocabulary allows to semantically annotate public transport entities with accessibility data that can be used to create richer applications on top of public transport data. The vocabulary extension is publicly available on Github⁶.

Linked Connections

The Linked Connections (LC) vocabulary⁷ describes the basic building blocks of the LC framework, called connections. A connection represents a specific travel, within a public transport network, made by a given vehicle and contains data about the departure stop, departure time, arrival stop and arrival time of that specific travel. Some of the properties defined by the LC vocabulary are the following:

- **lc:departureStop** - Represents the gtfs:stop from where a vehicle is departing.
- **lc:departureTime** - Contains the date and time at which a vehicle departs from the lc:departureStop.
- **lc:departureDelay** - Gives the amount of seconds a vehicle departure is estimated to be delayed from its original lc:departureTime.

This vocabulary provides a common semantic description of the behaviour of public transport networks through the LC framework, which is optimized for route planning applications. Using this vocabulary facilitates the development of multimodal route planners that take into account multiple public transport networks, as it reduces the complexity to integrate them through common semantics.

Mapping Tools

In order to facilitate the use of the above described semantic vocabularies, during the project we developed tools that allow to map public transport datasets formatted using the GTFS and GTFS-RT formats to Linked Data. The tools we created are the following:

- **gtfs2lc**: This tool is a Node.js application that takes a GTFS dataset and converts it into a Linked Connections graph, using the Linked GTFS and Linked Connections vocabularies. The source code of the tool is available in Github⁸.
- **gtfsrt2lc**: This tool is a Node.js application that converts GTFS-RT updates into Linked Connections by using both the Linked GTFS and Linked Connections semantic vocabularies. The source code of this application is available in Github⁹

⁶ <https://github.com/oSoc18Spain/onto-transacc/blob/master/TransportAccessibilityv2.owl>

⁷ <http://semweb.mmlab.be/ns/linkedconnections#>

⁸ <https://github.com/linkedconnections/gtfs2lc>

⁹ <https://github.com/linkedconnections/gtfsrt2lc>

Public Services Ontologies

In this section we present the selected vocabularies used on the AccessFlanders application to semantically describe public service datasets, specifically focusing in the accessibility of public buildings.

Core Public Service Vocabulary Application Profile

Created by the European commission ISA² program¹⁰, the Core Public Service Vocabulary¹¹ (CPSV) is a simplified, reusable and extensible data model that captures the fundamental characteristics of a service offered by public administration. It is designed to facilitate the exchange of basic information about the functions carried out by the public sector and the services in which those functions are carried out. Figure 1 presents the data model defined by CPSV.

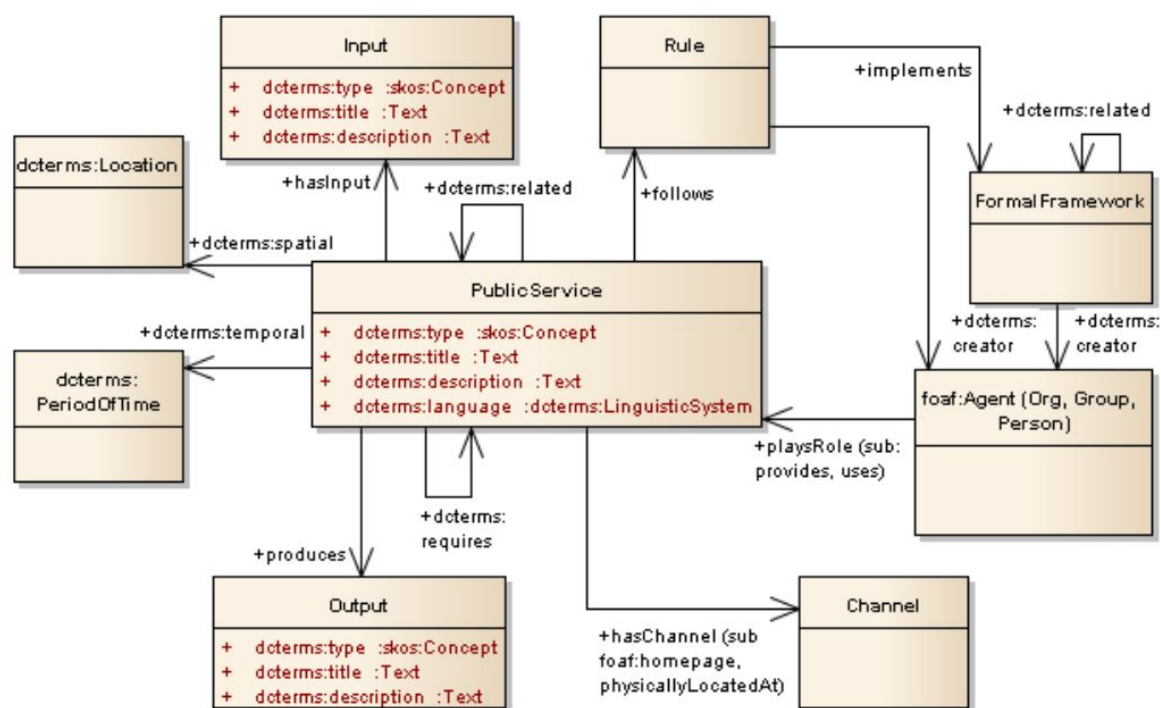


Figure 1. UML diagram for the Core Public Service Vocabulary¹².

¹⁰ https://ec.europa.eu/isa2/isa2_en

¹¹ https://joinup.ec.europa.eu/sites/default/files/distribution/2013-11/cpsv_101.ttl#

¹² <https://joinup.ec.europa.eu/release/core-public-service-vocabulary/101>

The main class defined by the CVSP vocabulary, as shown in figure 1, is the **cvsp:PublicService**. This class represents the service itself. A public service is the capacity to carry out a procedure and exists whether it is used or not. It is a set of deeds and acts performed by or on behalf of a public agency for the benefit of a citizen, a business or another public agency.

Taking the CVSP ontology as a core vocabulary, the CVSP-AP¹³ (Application-Profile) was defined. An Application Profile is a specification that reuses terms from one or more base standards, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used. The CVSP-AP focused ultimately on improving and harmonising the provision of information about public services on established eGovernment portals and adopting a user-centric perspective focusing on citizens and businesses. The CPSV-AP has been seen as a first step for creating a model for describing public services related to business and life events, to facilitate the set-up of catalogues of services oriented to businesses and citizens.

Infrastructure for Spatial Information in Europe

The Infrastructure for Spatial Information in Europe (INSPIRE) initiative aims at improving the sharing of spatial data and services between public authorities in Europe and in particular between the Member States and the European Institutions. INSPIRE addresses the interoperability of geospatial data sets and services for the exchange of data related to one of the 34 spatial data themes defined in the INSPIRE Directive¹⁴. Following this initiative, the European Commission Joint Research Center proposed a set of guidelines¹⁵ to publish the INSPIRE data models as Linked Data vocabularies through the RDF format. The published vocabularies are available on Github¹⁶ and give a semantic definition to concepts and properties defined on the INSPIRE schemas. Next there is an example of concept definitions taken from the ontology derived from INSPIRE Addresses application schema:

- **ad:Address** - An identification of the fixed location of property by means of a structured composition of geographic means and identifiers.
- **ad:PostalDescriptor.postCode** - A code created and maintained for postal purposes to identify a subdivision of addresses and postal delivery points.
- **ad:AdminUnitName** - An address component which represents the name of a unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional or national governance.

¹³ <https://joinup.ec.europa.eu/release/core-public-service-vocabulary-application-profile/21>

¹⁴ <https://inspire.ec.europa.eu/data-specifications/2892>

¹⁵ <http://inspire-eu-rdf.github.io/inspire-rdf-guidelines>

¹⁶ <https://github.com/inspire-eu-rdf/inspire-rdf-vocabularies>

On the above example the prefix **ad** stands as an abbreviation of the IRI [<http://inspire.ec.europa.eu/ont/ad#>](http://inspire.ec.europa.eu/ont/ad#).

OSLO Building Vocabulary

The OSLO (Open Standards for LOcal administrations) specification is the result of a public-private partnership initiated by V-ICT-OR¹⁷, the Flemish Organization for ICT in Local Government, and funded by Flemish ICT service providers and public administrations. OSLO aims to transform IT-service delivery efforts to focus on government customers: citizens and business enterprises. This Strategy focuses the government's energy on the use of technology to transform the delivery of services so that citizens and business only have to Ask Just Once to get what they need from their government. The OSLO Vocabulary is a simplified, reusable and extensible data model that captures the fundamental characteristics of information exchanged by public administration in the domains of contact information, localization and public services.

Extending the OSLO initiative, the government of Flanders proposed the OSLO Building vocabulary which focuses on terms that relate directly to buildings and building units. The Flemish Basic Registry¹⁸ was taken as a starting point on the definition of the vocabulary but also a coordination effort to align the vocabulary with the INSPIRE-Building application schema was made. The vocabulary has been publicly available¹⁹ by the Flemish government and provides definitions as follows:

- **gebouw:Gebouw** - A closed and/or covered above-ground or underground structure, which serves or is intended, either to house people, animals and objects, or to manufacture economic goods or to provide services.
- **gebouw:bouwjaar** - Date on which the building was completed.
- **gebouw:Gebouw geometrie** - 2D geometric representation of the building.

On the previous OSLO example the prefix **gebouw** stands as an abbreviation of the IRI [<http://data.vlaanderen.be/ns/gebouw#>](http://data.vlaanderen.be/ns/gebouw#) .

¹⁷ <http://www.v-ict-or.be/>

¹⁸

<https://overheid.vlaanderen.be/informatie-vlaanderen/wie-zijn-we/communicatiefiche-bouwsteen-vlaamse-basisregisters>

¹⁹ <https://data.vlaanderen.be/ns/gebouw>

Wheelchair Accessibility Vocabulary

The Wheelchair accessibility is an RDF vocabulary²⁰ defined to describe wheelchair accessibility potential of building related spaces. This vocabulary provides a set of properties to physically describe whether a space in a building (e.g. entrance, elevator, stairwell, toilet, etc.) is adequate for wheelchairs and their users. Among the properties defined in the vocabulary are the following:

- **wa:elevatorDoorWidth** - Provides the width in cm of an elevator door.
- **wa:toiletSteepness** - Measures the slope on the route towards a given toilet.
- **wa:handrails** - Indicates if there are handrails available in a certain place.

The **wa** prefix on the previous example stands as an abbreviation of the IRI <http://semweb.mmlab.be/ns/wa#>.

²⁰ <http://semweb.mmlab.be/ns/wa#>