Journal of Universal Computer Science, vol. 21, no. 6 (2015), 777-795
submitted: 15/11/14, accepted: 20/2/15, appeared: 1/6/15 © J.UCS

Modelling and Linking Accessibility Data in the Public Bus Network

Paloma Cáceres, Almudena Sierra-Alonso, Carlos E. Cuesta, Belén Vela
José María Cavero
(Universidad Rey Juan Carlos, Madrid, Spain
{paloma.caceres, almudena.sierra, carlos.cuesta, belen.vela, josemaria.cavero}@urjc.es)

Abstract: Many organizations and public administrations are currently working to open their data for the general use, in the context of the Linked Open Data (LOD) initiative. This emphasizes the need to publish available data in a structured format, so that they can be accessed and externally processed in order to maximize their utility for citizens. This also poses a general challenge in the area of information technologies, as it requires being able to integrate data from heterogeneous distributed sources, provided in a compatible format. We consider the specific domain of public transport networks, where the information has obviously a public interest. The processing of these data can be used for a wide variety of applications: route planning systems are a classic example. However, most of the existing approaches in this domain fail to provide specific means to deal with *accessibility data*, i.e. the information which is relevant for people with special mobility needs, especially when it is obtained from different transit networks. In this context, the LD approach is particularly useful, due to the complexity of those relationships and their inherently graph-oriented nature. In this work we describe the process we use to define this information and to make it available in the context of the CoMobility project. First, we define a conceptual model, supported by the main reference data models in the transport domain: Transmodel and IFOPT, and emphasizing the role of accessibility. We transform it into an ontology, in order to combine data from diverse sources; and then such concrete data are captured using the schema defined by this ontology. In this context, we use actual data from the public bus network in the City of Madrid. Afterwards we are able to define and build new mobility services, e.g. to automatically decide if a certain bus route is *accessible* or not, considering any potential transfers in this route. Finally, we are able to publish the obtained information, albeit partially, preserving the LOD spirit.

Keywords: Open data, semantic web, public transport networks, Transmodel, IFOPT, RDF
Categories: H.3.0, H.4.0, I.2.4, K.4.2, M.6

1 Introduction

Currently, the Linked Open Data (LOD) [LOD, 14] initiative presents a challenge in the area of Information Technologies (IT). It provides the mechanism for publishing, enriching and sharing data, information and knowledge on the Web, using *semantic web* technologies. It comprises the following principles: (a) using Universal Resource Identifiers (URIs) [URI, 14] for identifying all kinds of elements ("things"), (b) making these URIs accessible via the HTTP protocol and (c) providing a description for these things using the Resource Description Format (RDF) [Klyne, 12] along with (d) URI links to related information.

This paradigm can be applied to heterogeneous distributed data sources, which are therefore integrated, and published as LOD. This way, a particular application is now able to use the information of public transport networks, which of course must be considered of public interest. Due to the large size and variety of these data, LOD provides both the mechanism to publish them and to do it in a flexible way, able to support the definition of new mobility services for citizens. A classic example could be route planning systems, which combine data from different public transit networks. The relevance of open data in the context of transport networks is significant, and this can be shown using both research papers and communications from several big organizations, for instance [Epsi, 12], [Hobbs, 14], [ODUK, 14].

Having this in mind, we intend to apply Information Technologies to the task of improving mobility services – specifically considering the case of citizens, both in usual or occasional trips, and trying to optimize their intended routes by using any available means of transport: both public transport and the rational sharing of private transport. Therefore we have designed an IT platform, called CoMobility [Cuesta, 13], to assist those citizens in the use of intermodal transport sharing, and integrating *carpooling* with public transport, as well as other private transport media. Besides, we also consider the *accessibility* of these media as a relevant aspect of public transit networks, as it makes possible the mobility of people with special needs. We want to emphasize that, in this article, the term "*accessibility*" is specifically referring to "accessibility for people with special needs" (that is, accessibility for blind people, or people using a wheelchair, who have permanent mobility issues; but also people pushing a baby carriage or carrying luggage, etc).

Which data is required from transport networks, when users in transit make a request to a service to define a route? Is there any support information about the transport network already available? Trying to answer these questions, we have studied the already existing transport metamodels, specifically designed to be generic: Transmodel [Transmodel, 14] and IFOPT [IFOPT, 14]. Our intention was to match the actual data from public transport networks with such standards. Transmodel is a European Reference Data Model for Public Transport Information, which provides a model of public transport concepts, and data structures that can be useful to build information systems related to the different kinds of public transport. But Transmodel does not provide relevant information about *accessibility*. For this reason, we have also studied the related IFOPT metamodel, conceived itself as an extension of Transmodel. It defines a model (and also the identification principles) for the main *fixed objects* related to public access to Public Transport (e.g. stop points, stop areas, stations, connection links, entrances, etc.). It already includes specific structures to describe accessibility data about the equipment of vehicles, stop places and access areas. Therefore, we have defined a conceptual model by means a UML class diagram, supported by the IFOPT model as a basis, which includes new (additional) accessibility elements (classes and attributes) related to vehicle equipment.

This model describes our *reference ontology*, designed with the purpose to define *accessibility features* in public transport, and specifically on buses. There are already several existing works dealing with ontologies for public transport in the literature. [Timpf, 02] proposes an ontology of wayfindings, from a traveller's perspective. His work is not based on any transport metamodel or standard, and it does not take into account any information about accessibility elements of public transport. The same happens with [Becker, 97], which describes an ontology for public transport which is actually based on a pre-existing, generic ontology for scheduling; it lacks a lot of detail about the transport domain, and does not consider accessibility issues. Other proposals define more evolved ontologies for public transport, either following or considering Transmodel: [Houda, 10], [Marçal de Oliveira, 13]. But again, like in the works mentioned above, they do not include any information about accessibility elements in public transport.

To simplify the exposition, this article focuses on a specific issue in accessibility, namely how to publish the information about the equipment of public buses and their accessibility, as open data. However the intention is not to lose any generality: only the presentation is simplified – our models, ontology and applications still consider many relevant aspects of accessibility, besides equipment itself.

The format of public data, within many existing open data initiatives, prevents non-experts from using them directly, and thus it requires additional semantics, as provided by Linked Open Data [LOD, 14], [Heath, 11]. There are several proposals in this regard, within the specific context of transport networks. [Colpaert, 14] presents an immature proposal of a route planning system which uses Linked Open Data, as an initial idea of a doctoral thesis; but it never takes into account accessibility elements. [Pham and Jung, 14] presents a workflow for publishing and linking transport data on the web; moreover, they apply their linked transport data proposal in two different real-world datasets. But again, they do not take into account accessibility issues, either.

In this article, our proposal focuses on modelling the accessibility and transport data for the public bus network. To do so, we take the following steps: first, we analyze the format of original data format, and identify the data semantics. Second, we match these data against the vocabulary included in standard transport metamodels, both Transmodel and IFOPT, and we add relevant information when necessary, enriching those models. And third, we define an extended ontology as an RDF Schema, highlighting the details about the equipment of public buses and ultimately representing and providing them as Linked Open Data.

The paper is structured as follows: in section 2, we briefly introduce the context of this work: the CoMobility project. Section 3 describes the technologies we use in the Linked Open Data context: the Resource Description Framework (RDF) and the graph-oriented Query Language (SparQL) for RDF. In section 4, we define a specific process to publish accessibility elements and data of transport media as LOD. To illustrate this process, the section 5 presents a case study describing the publication of accessibility data – the case focuses on the support for wheelchairs in public buses of the EMT transport company, in the City of Madrid. Additionally, section 6 describes a Jena-based web user interface intended to show the structure of our proposal by exposing our RDF Schema and a partial RDF dataset. In section 7 we summarize the main conclusions of our work; and finally, several paths for future work are briefly discussed in section 8.

2 Context: the CoMobility Project

The CoMobility project [Cuesta, 13] defines a multimodal architecture based on linked open data for a sustainable mobility. Its main goals are improving the citizen mobility, optimizing their trips combining both public transport and private sharing transport (i.e. car sharing), providing accessible trips when necessary and saving energy and reducing the pollution.

We have developed a systematic approach to (i) accessing open, integrated and semantically annotated transportation data and street maps, (ii) combining them with private data, and (iii) supplying mechanisms to allow the actors to share and search these data. Therefore, its conceptual architecture provides the means to perform the following tasks: First, the platform can identify, select, extract and integrate data from different and heterogeneous sources, stemming from the transportation, geographical and energy domains. Second, data from public institutions are obtained automatically in the form of open data. Third, these data are annotated as linked data, and a set of heuristics generate links between data items from different sources without human intervention. Fourth, these data are integrated with private data provided by users themselves. And finally, CoMobility provides intuitive and customized data analytics and visualization, allowing individuals to become aware of the environmental impact of their transport choices. Figure 1 provides a general depiction of our project.

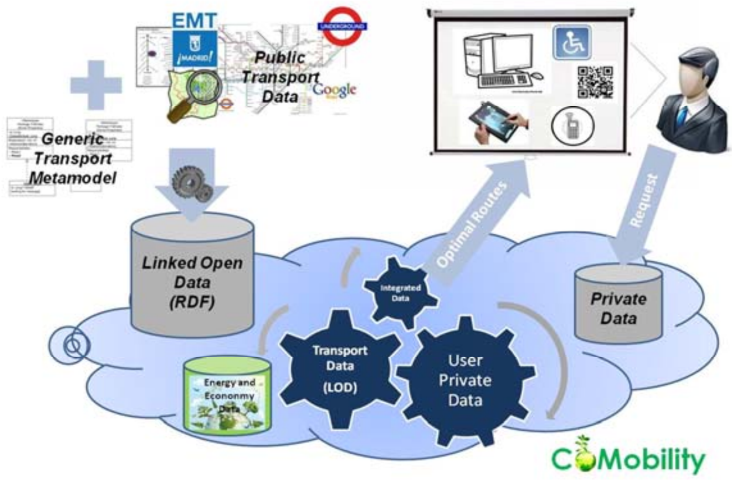

Figure 1: CoMobility project

This work is focused on public transport data, the generic transport metamodels and Linked Open Data (in RDF). In section 4, we describe in depth the process combining all of them in the context of accessibility.

The CoMobility Project is supported by the Spanish Ministry of Economy and Competitiveness and several companies have expressed their interest in our results. The most representative are: the public bus company of Madrid (EMT Madrid) [EMTMadrid, 12], the Public Consortium for Transport in Madrid (CTM) [CTM, 14], and the Spanish National Society of Blind People (ONCE) [ONCE, 14]; as well as the Chair of EcoTransport, Technology and Mobility of the Rey Juan Carlos University [ChETM, 14], which also supports this work.

3 Technologies for Linked Open Data

In this section we introduce the technologies in the open data area used in this work: RDF and SPARQL.

3.1 A brief introduction to RDF

The Resource Description Framework (RDF) [Klyne, 12] provides an extremely simple data model in which entities (also called resources) are described in the form of triples (subject, predicate, object).

RDF has been gaining momentum since its inception thanks to its adoption in diverse fields, such as bioinformatics, social networks, or geographical data. The Linked Open Data project plays a crucial role in the RDF evolution [LOD, 14]. It leverages the Web infrastructure to encourage the publication of such semantic data, providing global identity to resources using HTTP URIs. Moreover, integration between data sources is done at the most basic level of triples, that is, to connect two data sources can be as easy as making connection between those resources.

This philosophy pushes the traditional document-centric perspective of the Web to a data-centric view, emerging a huge interconnected cloud of data-to-data hyperlinks: the Web of Data. Latest statistics point that more than 31 billion triples were published and more than 500 million links established cross-relations between datasets.

Although each piece of information could be particularly small (the Big Data's long tail), the integration within a subpart of this Web of Data can also be seen as an example of Big Semantic Data.

It is worth noting that, RFID labels, Web processes (crawlers, search engines, recommendation systems), *smartphones* and sensors are potential sources of RDF data, such as in our previous use cases. These are the main players in the so-called Internet of Things, in which the Linked Data philosophy can be applied naturally by simply assigning URIs to the real-world things producing RDF data about them via Web. As a result, the activity of all involved devices is recorded and linked between them, enabling large projects (such as the emergent concept of smart-cities) to be success-fully implemented.

3.2 SPARQL: a Graph-oriented Query Language for RDF

SPARQL is the W3C Recommendation [SparQL, 14] for querying RDF. It is a graph-oriented language regarding triple patterns as the atom queries. A triple pattern follows the RDF triple structure, but the subject, the predicate or the object may be a variable. These queries, generically referred as Basic Graph Patterns (BGPs), can be then combined through unions or optional constructions. Besides, the returned values can be filtered using different functions. Thus, SPARQL engines must provide, at least, fast triple pattern resolution and efficient join methods. Moreover, BGP resolution relies on obtaining effective execution plans which optimize the order in which each triple pattern is resolved. Although query optimization is orthogonal to the current work, it is worth noting that the architecture must provide effective mechanisms for its implementation.

4 The process to publish the accessibility data as LOD

This work focuses on modelling accessibility data for the public bus network using Linked Open Data (LOD). We would like to emphasize two features of this work; first, that we have worked with actual accessibility information and data from the Madrid public bus network, provided by EMT Madrid (the public bus company) [EMTMadrid, 12]. Second, that since its conception, we have followed the IFOPT metamodel [IFOPT, 14], the current standard to model the features of transport media, including accessibility. IFOPT defines a model and the identification principles for the main fixed objects related to the access to Public Transport (e.g. stop points, stations, stop areas, connection links, entrances, etc.). In particular, our work in this paper focuses specifically on a subset of the Stop Places model, namely the Vehicle Equipment submodel. Therefore, for the remainder of this paper, we will focus in the part of vehicle equipment which models accessibility features.

To achieve this goal, we have defined a stepwise refinement process, which we briefly describe in the following:

- First, we have studied the accessibility features of the public bus network, and the format in which the original data are provided. Later, we have identified the data semantics of the accessibility information;
- Second, we have analyzed the accessibility information included in the original IFOPT metamodel, and we have matched the original data of EMT Madrid against the vocabulary of this metamodel;
- Third, using the IFOPT metamodel as a basis, we have defined a conceptual model that describes the accessibility features of the public bus network, specifically in the domain of vehicle equipment. The result is an extension of IFOPT metamodel;
- Then, considering both this ontology and the available information, we have semi-automatically generated a RDF Schema [RDFS, 14] for this ontology;
- After this schema is completed, we have distilled the set of RDF data, by instantiating the concepts in our RDF Schema with real data about specific vehicles from the EMT Madrid network;
- Finally we have published this RDF dataset as Open Data, in a format which can be retrieved using SPARQL [SparQL, 14].

The first three points imply to analyze the scattered information and the IFOPT standard to conclude in a new model to add the new discovered elements to the metamodel. The rest of activities are directed to implement and to publish the data as open linked data. In the Figure 2, that shows the portion of the CoMobility project in which our work is focused, we can see these two objectives. On the top, the Public Transport Data plus the Generic Transport Metamodel represent the analysis activities to achieve a new metamodel which includes the accessibility features. The grey arrow symbolizes the process of transformation from this metamodel into a RDF Schema and the final data publication according that schema.

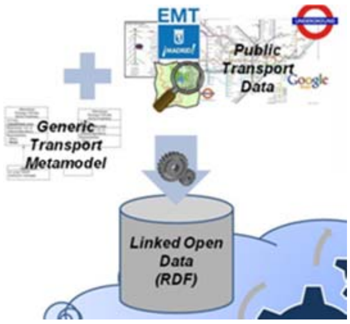

Figure 2: Our focus in this article

Following subsections are focused on the process to achieve the open data final model about accessibility information of public transport buses: section 4.1 describes the modifications to the IFOPT metamodel to define a new accessibility model and section 4.2 shows the generation of RDF Schema preserving the knowledge present in that metamodel and the accessibility data publication.

4.1 A Conceptual Model for Accessibility

The accesibility information for buses, as originally presented, is scattered across several documents (text, word or pdf) and web pages. Some of them are public, such as [CTM, 13], [IUEE, 02] or [Vega, 06]. Other documents were directly provided by EMT Madrid. So, to extract the information related to accessibility in buses, we had first to study carefully the available documents. Table 1 shows one of the first results of this study: a list of accesibility elements that can be found in buses.

The next step has been to study the transport standard models Transmodel (V5.0) [Transmodel, 14] and IFOPT [IFOPT, 14], to match them against the bus accessibility data from EMT Madrid. As already noted, Transmodel is the European Reference Data Model for Public Transport Information, which provides a model of public transport concepts, and data structures that can be used to build information systems dealing with different kinds of public transport. But Transmodel does not describe the accessibility features of the different transport media. These features are described and modelled using the IFOPT metamodel, conceived itself as an extension of Transmodel. It defines a model for the main *fixed objects* related to public access to Public Transport (e.g. stop points, stop areas, stations, connection links, entrances, vehicle equipment, etc.), as well as the identification principles for those fixed points. As also noted, our work in this article focuses on the Vehicle Equipment sub-model, which is a part of the Stop Places model of IFOPT. We specifically focus in the part of Vehicle Equipment which models accessibility features. The IFOPT metamodel, in order to model accessibility features, defines the following classes, namely `WheelchairVehicleEquipment`, `AccessVehicleEquipment` and `VehiclePassengerInfoEquipment`.

ACCESSIBILITY ELEMENTS
Kneelling or neumatic lowering
Ramp
Buttom request ramp
Low floor
Ramp bearing capacity
Automatic doors
High Contrast buttons
Buttons at low height
Contrast band doors
Antiskid floor
Ilumination
Enlarged Central Platform
Paddest Back Rest
Vertical rods
Handrails
Tactile guidance steps
Back bicycle support
Baby chairs
Special seats
Space for baby carriage
Space for bagagge
Lugagge belt
Slip band for lugagge
Audio and visual signals for stops

Table 1: Accessibility elements in buses.

Figure 3 pictures these classes with a white background. Their meaning is:

- The class `WheelchairVehicleEquipment` models the equipment designed for wheelchairs.
- The class `AccessVehicleEquipment` designates elements to facilitate the access to the vehicle. Some attributes of this class, such as `LowFloor` or `Ramp,` are specifically conceived for people with special needs.
- The class `VehiclePassengerInfoEquipment` models the facilities inside the vehicle. To designate the facilities for special needs the attribute `AccessibilityInfoType` is defined. The values of this attribute are in turn modelled as another class: `VehicleAccessibilityInfoEnum`. This class models, for example, special formats for messages: `audioForHearingImpaired` or `displaysForVisuallyImpaired`.

To model EMT buses accessibility data using the Vehicle Equipment in IFOPT, we first tried to instantiate this model and to populate it using EMT accessibility data. As result, we have detected that the original, “vanilla” IFOPT model does not completely fit our purpose, as there are accessibility elements in EMT buses that we are not able to model using IFOPT Vehicle Equipment.

When, for example, we want to know if a bus has special seats, and the width of those seats, we do not have attributes in IFOPT (nor in Transmodel) to define these data. So we add the class `SpecialVehicleEquipment`, with the attributes `SpecialSeats` and `SpecialSeatWidth` to indicate if the bus has special seats (or not) and their width; the attribute `NumberSeatReserve` indicates whether there are reserved seats to people with special needs or not. This class also contains attributes to determine if the bus has a baby chair, space for baggage or baby carriages, belt or split band for luggage. We have designed four other additional UML classes to model accessibility elements presents in EMT buses. They are: `HandrailsVehicleEquipment,` to describe the handrails in the access area; `BicyclesEquipmentVehicle,` to determine whether the bus has equipment to carry bicycles; `CentralPlatformVehicleEquipment,` to specify if the vehicle has special elements in that platform, such as a button to request the ramp; and `StopRequestVehicleEquipment,` to indicate if there are special elements to help people to request the bus to stop.

We have also added extra attributes in existing classes. For example, the `ButtonRequestRamp` attribute has been added to the `AccessVehicleEquipment` class. It has the purpose is to indicate if there is a button outside the bus to request the ramp to be deployed. Another new attribute is `NeumaticLowering` (or kneeling) to register this facility.

We can see the result of these modifications in Figure 3. As noted, we simply enrich the original model. The new classes and attributes are shaded in blue.

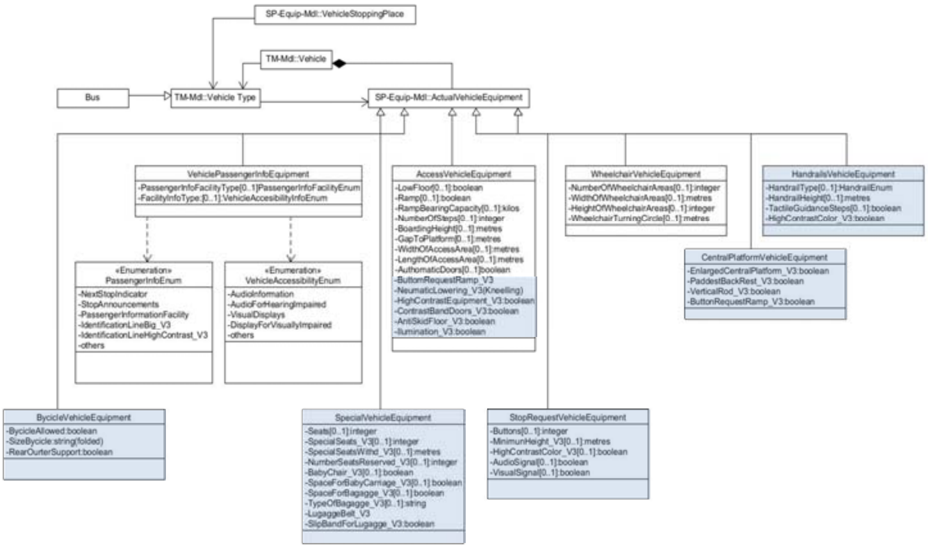

Figure 3: The IFOPT model, extended to include the accessibility features in EMT buses.

4.2 Converting Accessibility Information into LOD

As said before, the next step to publish accessibility data for public buses as open data is to generate an adequate RDF Schema, able to capture the ontology corresponding to the model in the previous section. This schema must preserve the structure of the modified IPOFT model, as it is the recognized standard in the transport area.

Starting from the conceptual model (see Figure 3), we have defined the RDF Schema that describes these accessibility features. This schema has been instrumented using Apache Jena [Jung 12] within Eclipse [Eclipse, 14]. Jena is a free and open source Java framework for building Semantic Web and Linked Data applications.

```
String SOURCE = "http://www.vortic3.com/IFOPT";
String NS = SOURCE + "#";
Stringlang = "en";

// Create an empty ontology model

OntModelontologia = ModelFactory
                      .createOntologyModel(OntModelSpec.RDFS_MEM);
ontologia.setNsPrefix("v3", NS);

// OntologyClasses

OntClassvehicle_class = ontologia.createClass(NS + "");
OntClassbus_class = ontologia.createClass Vehicle (NS + "Bus");

vehicle_class.addSubClass(bus_class);
vehicle_class.addSuperClass(RDFS.Resource);

OntClassactualEquipmente_class= ontologia.createClass(NS
                             + "ActualEquipment");
OntClasswheelchair_class = ontologia.createClass(NS
                             + "WheelchairVehicleEquipment");
actualEquipmente_class.addSubClass(wheelchair_class);
actualEquipmente_class.addSuperClass(RDFS.Resource);

// OntologyProperties

OntPropertyhasEquipment_property = ontologia.createOntProperty(NS
                             + "hasEquipment");
hasEquipment_property.addProperty(RDFS.comment,
                             "Actual equipment of vehicle", lang);
hasEquipment_property.addProperty(RDFS.domain, vehicle_class);
hasEquipment_property.addProperty(RDFS.range, actualEquipmente_class);
```

Figure 4: Fragment of Jena code for the vehicle ontology.

Figure 4 shows a fragment of the Jena code. The classes which model the vehicle and its equipment, as originally described in the UML diagram, are now defined as ontology classes in the RDF schema. Then, subclass and superclass relationships between them are also modelled and established. Next, the composition relationship between vehicle and actual equipment classes is specified as an ontology property. The specific UML class for wheelchair equipment is first created as a class of the ontology, and then it is defined as subclass of actual equipment of the vehicle. The wheelchair equipment has been included to show how to describe this kind of equipment. The rest of equipment elements are defined in a similar way.

From the code, a RDF schema graph [RDFS 14] has been generated, using a semi-automatic process which involves both Jena [Jung 12] and Protegé [Protegé, 14]. Figure 5 shows the RDF scheme graph obtained from the execution of Jena code. We want to check whether the UML diagram structure is preserved in the graph: the UML classes `Vehicle`, `ActualVehicleEquipment`, `Bus` and `WheelchairVehicleEquipment` are represented by `v3:Vehicle`, `v3:ActualEquipment`, `v3:Bus` and `v3:WheelchairVehicleEquipment` and defined as `rdfs:type` of `rdfs:Class` to maintain the conceptual structure. Furthermore it is established that `v3:Bus` and `v3:WheelchairVehicleEquipment` are subclasses of (i.e., predicate `rdfs:subClassOf`) `v3:Vehicle` and `v3:ActualEquipment`, respectively, just like in the UML class diagram.

The relationship between the vehicle (UML class `Vehicle`) and its specific equipment (UML class `ActualVehicleEquipment`) is defined by means of the schema property `v3:hasEquipment`. This property is added as domain of `v3:Vehicle` and as range of `v3:ActualEquipment`.

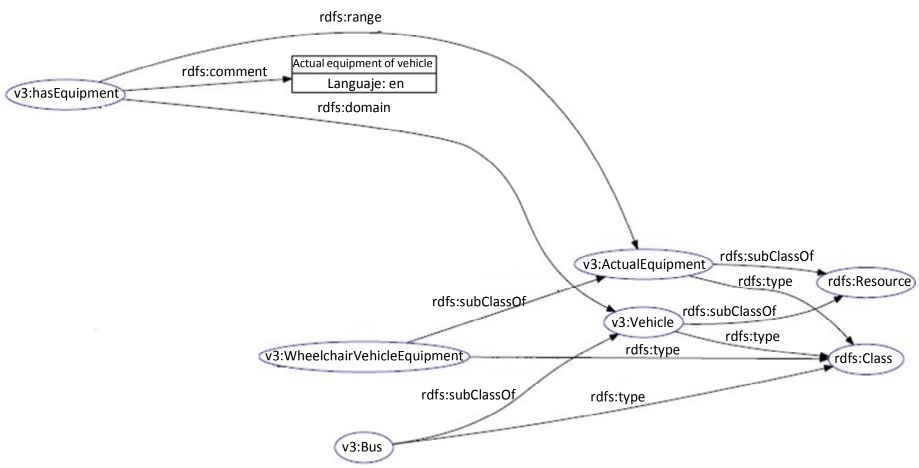

Figure 5: Fragment RDF Scheme for vehicle ontology.

The next step of the process would be to provide a full RDF dataset, including specific data for these bus accessibility features. EMT Madrid has provided us with real data about these features. Once obtained, this RDF dataset has to be validated against the RDF Schema. The compliance of the dataset with the RDF schema has been automatically checked using Jena, again.

In the next section, a complete case study is described following this process; we also provide a partial view of the RDF dataset.

5 Case Study: providing wheelchair accessibility information for the Madrid public bus network as LOD

In this section we present a case study to illustrate the process described in the previous section, which transforms bus accessibility data from the original sources into an open semantic format.

After studying the accessibility features for the public bus network, we have analyzed the data semantics for this accessibility information. We have found that some vehicles have a specific area for wheelchairs, in which both the height and the width of access area, and the turning circle area (i.e. the space to turn the wheelchair) are also specified. Then we have compared this information to the accessibility elements in the IFOPT metamodel. Specifically, IFOPT provides a class for this purpose, namely the `WheelChairVehicleEquipment` class, in its UML version, which specifies the following attributes:

- `NumberOfWheelChairAreas[0..1]:integer`
- `WidthOfAccessArea[0..1]:metres`
- `HeightOfAccessArea[0..1]:integer`
- `WheelChairTurnningCircle[0..1]:metres`

As we can see, IFOPT supports all the required information about wheelchairs for the current network of public buses from EMT Madrid. Then, considering this information, we have been able to define an ontology for it, describing the vehicle equipment, and specifically focusing on the accessibility features for public buses. This ontology summarizes the knowledge contained in the UML version of the IFOPT sub-model *Vehicle Equipment*, as already indicated. The structure of this class diagram has been directly instrumented into Apache Jena.

Figure 6 shows a fragment of the Jena code. This code fragment defines the attributes of the wheelchair equipment adding them to `wheelchair_class` domain.

```
// Specific properties of wheelchair equipment

OntProperty numberOfWheelchairAread = ontologia.createOntProperty(NS
                              + "numberOfWheelchairAreas");
numberOfWheelchairAread.addRange(XSD.xint);
numberOfWheelchairAread.addDomain(wheelchair_class);
OntProperty wheelchairTurnningCircle = ontologia.createOntProperty(NS
                              + "wheelchairTurnningCircle");
wheelchairTurnningCircle.addRange(XSD.xdouble);
wheelchairTurnningCircle.addDomain(wheelchair_class);
OntProperty heightOfAccessArea = ontologia.createOntProperty(NS
                              + "heightOfAccessArea");
heightOfAccessArea.addRange(XSD.xdouble);
heightOfAccessArea.addDomain(wheelchair_class);
OntProperty widthOfAccessArea = ontologia.createOntProperty(NS
                              + "widthOfAccessArea");
widthOfAccessArea.addRange(XSD.xdouble);
widthOfAccessArea.addDomain(wheelchair_class);
```

Figure 6: Fragment of Jena code for the wheelchair equipment.

This code fragment, added to the previous fragment (see Figure 4), generates the RDF Schema [RDFS 14] for the wheelchair equipment. Figure 7 shows a part of that RDF schema, dealing with the vehicle equipment for wheelchair. We can see how the information of the UML class diagram is maintained. In the UML class diagram (see Figure 3), the `WheelchairVehicleEquipment` class has four attributes, which are modelled in the RDF Scheme with `rdf:domain` predicates. Furthermore, it is also indicated that this class is a subclass of the class `ActualVehicleEquipment` by means of the predicate `rdf:subClassOf` which has the subject `v3:ActualEquipment`.

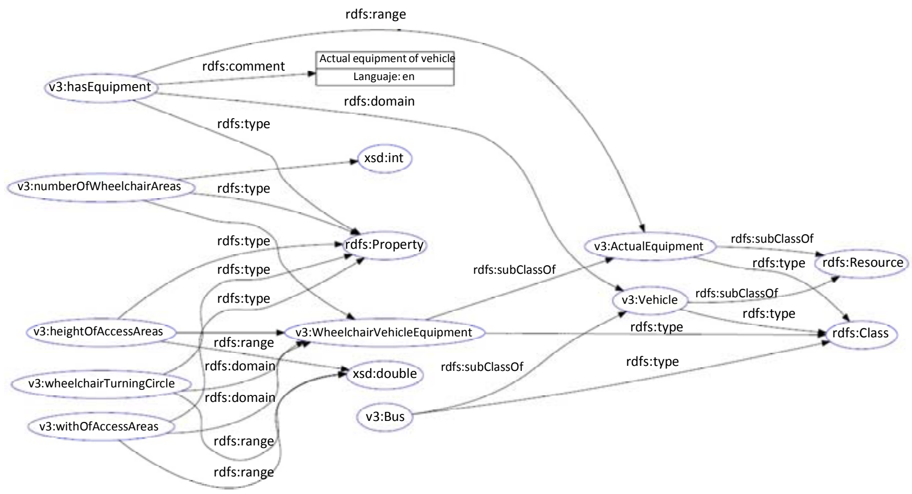

Figure 7: Fragment of the RDF Schema for the wheelchair ontology.

In the last step of the process, we have generated the RDF dataset for this information, using real-world data provided by EMT Madrid. According to these data, a specific bus (`Bus_1`, to simplify the identification) has two Wheelchair Areas. Each one of them has an access area width of 1.6 metres, an access height area equal to 2.6 metres, and the turning circle area is 3.15 metres. The code to generate the instance with these data is presented in Figure 8. The RDF data diagram for this information is provided as Figure 9.

As we can see, the RDF data tells us that the bus *Bus_1* has equipment for 2 wheelchairs (predicate `v3:numberOfWheelchairAreas`) and the measures of access area (predicates `v3:heightOfAccessArea`, `v3:widthOfAccessArea` and `v3:wheelchairTurnningCircle`).

Once obtained and fully populated, this model has been checked against the previous RDF Schema. Therefore, the RDF dataset has been shown to be fully compliant to that RDFS, using Jena for that purpose.

After this verification, we are finally able to upload the full dataset, including this accessibility information. We could use any RDF store to make this dataset available; for convenience, we have used Fuseki [Fuseki, 14], another element of the Apache

Jena Project. Fuseki is a natural choice, as it already provides the required capabilities for querying this information, using SPARQL; however, currently we only use this platform within prototype applications. We also provide a Jena-based programming interface to access the same information, mostly to perform very basic queries and to be able to directly manage these data. Additionally, as we describe in the next section, we also use this Jena programming interface to provide a simple web-based user interface, with the purpose of exposing at least a significant part of the stored information, in accordance with the LOD spirit.

```
// Instance

OntModel ontData=ModelFactory
        .createOntologyModel(OntModelSpec.RDFS_MEM);
ontData.setNsPrefix("v3", NS);
 Individual instance_bus =
    ontData.createIndividual(bus_class.toString() + "_1", bus_class);
instance_bus.addLabel("Bus_1", lang);
Individual instance_wheelChair =
   ontData.createIndividual(wheelchair_class.toString()+"_standard",
   wheelchair_class);
instance_wheelChair.addLiteral(numberOfWheelchairAreaд, 2);
instance_wheelChair.addLiteral(wheelchairTurnningCircle, 3.15);
instance_wheelChair.addLiteral(heightOfAccessArea, 2.60);
instance_wheelChair.addLiteral(heightOfAccessArea, 1.60);
instance_bus.addProperty(hasEquipment_property, instance_wheelChair);
```

Figure 8: Fragment of Jena code to create a wheelChair instance

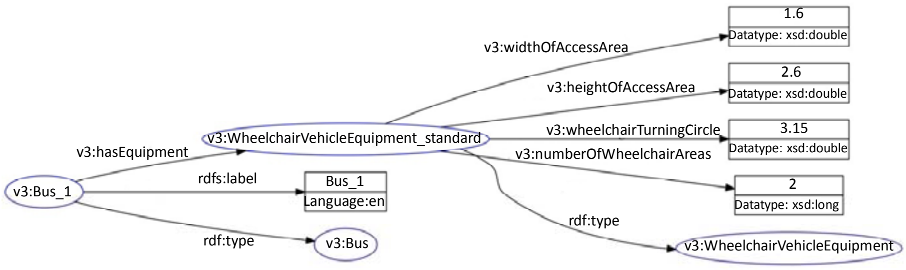

Figure 9: RDF data diagram for wheelchair equipment of the bus Bus_1.

6 Exhibit: A Jena-based web user interface

Even though the whole system has been designed with specific functionalities in mind – to assist the user in the transport network, including accessibility issues, it has also been designed with an open data philosophy, i.e. everything that can be made open, is exposed as open. This is conceived as a general strategy, not only for reasons of social utility, but also because it simplifies further extensions of the system, and even the design of any additional application.

However, we have also to comply with non-disclosure agreements (NDA) from our data providers – that means that though we are authorized to use these data and even to disseminate the results of their application, we are not able to provide the original data either in raw or pre-processed format.

Therefore, we are not able to simply provide an open interface to our data, and most of the complexity of the system gets encapsulated within our applications. However, having the open data spirit in mind, and trying to find a compromise, we intend to provide an additional, more generic interface for these data. The main purpose of this interface is to give a general impression of the information which is stored in our semantic database. Hence we provide a very simple, and limited, Jena-based web interface capable to access that knowledge. The purpose of that interface is mainly illustrative.

This web interface is available in http://193.147.62.170:8080/EMTProject/. It provides access to five different perspectives of the data stored in our Jena triple-store. Figure 10 shows the home page of the web user interface.

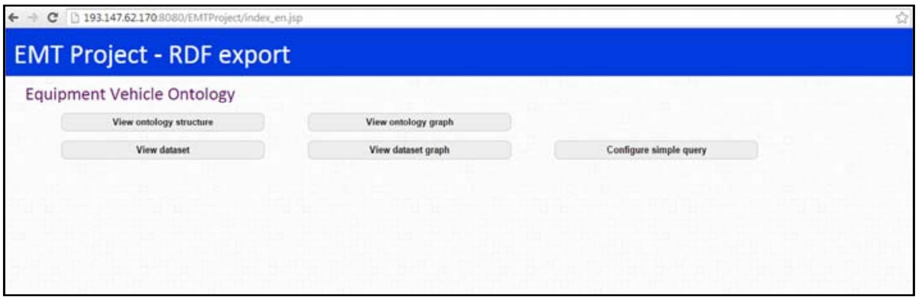

Figure 10: Index page for our web-based user interface

The first one is the full definition of our RDF Schema, directly retrieved from our Jena datastore – this RFDS is therefore provided as a sequence of NTriples in tabular form, which is actually used to check the well-formedness of our data. The second perspective consists of a graphical depiction of the same ontology, which was originally generated for the RDFS definition. These structures can be disseminated in their complete form, as they have been generated from our domain model, and they do not imply any violations of the NDA.

The third perspective is a partial presentation of some of the data themselves, which comply with this schema definition. The purpose is just to provide an impression of the kind of data which are available – hence it is not by any means complete, but just a small part of the stored data, to avoid any accidental disclosure. The RDF triples are provided, again, as NTriples in tabular form. Figure 11 shows a partial view of the dataset. At the bottom of this figure we can, for example, see that the `numberOfWheelchairArea` (Predicate) is 2 (Object) as a part of the `WheelchairEquipmentVehicle` (Subject).

Similarly, the fourth perspective provides a graphical depiction of this partial dataset. Finally, the fifth section of the web interface provides a very limited HTML form, which is able to forward very simple SPARQL queries to the Jena database, basically by fixing just one variable (choosing from subject, object or predicate). This interface does not allow the definition of complex queries, but it can be used to provide a more dynamic (i.e. interactive) impression of the information contained in the datastore, and the way in which certain triples are linked to other.

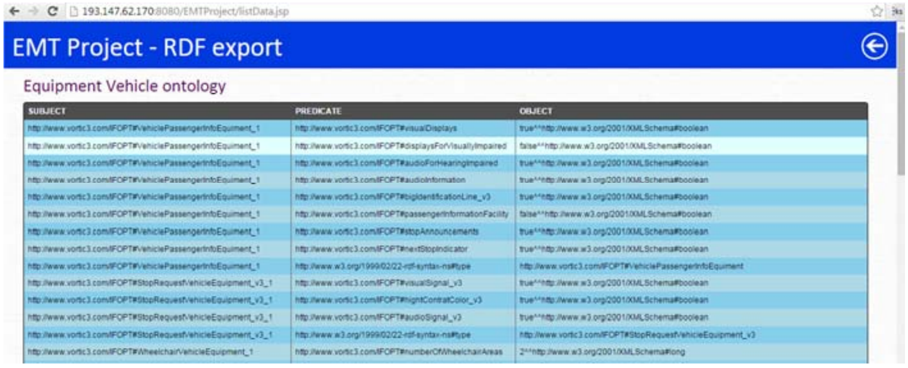

Figure 11: Partial view of the dataset, within the web user interface

Our semantic data are used in quite more complex ways (and generate much more complex SPARQL queries) within our system, and in particular within the Notify.me mobile application, which is able, e.g. to inform about the accessibility of a certain bus line at a concrete time – but these are conceived as specific applications. The web interface described here just intends that a technical user is able to have a limited interaction to the system, to acquire a general impression about the contained information, without violating the terms of our NDA.

7 Conclusions

In this paper, we present a process to define *accessibility data* in the context of public transport networks, and to publish them as part of the LOD initiative, specifically in a graph-oriented format – i.e. RDF schemas and datasets. As already noted, information about transport and transport networks, is always of a public interest, and hence it is particularly adequate to be exported in the LOD context – the publication of these data provides the opportunity to offer *new services* for public transport users, even in previously unforeseen situations.

We emphasize the specific case of accessibility for three reasons: first, to the best of our knowledge, there are very few examples of public transport networks where accessibility information is available as *open* data, even when it is critical for people with special mobility needs. Second, even when accessibility information is present, it usually refers to just one network, using some specific structure, and it does not use any kind of standard – neither a transport-based standard, nor a Linked Data standard which makes simpler to combine different sources. Our work using both RDF and an IFOPT-based model is novel in this regard. Third, once this information is available, we are able to develop new services: our initial case, though not the only one, consists of using a planning system to propose a route within the transport network, and to use these data to check the accessibility features of the proposed route. This is not a trivial challenge: once the user makes a request, the route must be composed of different sub-routes and transfers, each with their accessibility features; moreover, the question includes even concurrent issues – two subroutes might be accessible only partially, and their composition would only be accessible if it is adequately synchronized.

In this article we have presented a process to transform accessibility data from a public transport network (specifically, the public bus network of the City of Madrid) into a usable LOD format. To do so, we have had to characterize the data from the original source (EMT Madrid), and also the most adequate existing transport standard, i.e. IFOPT. This is one of the major novelties of this work – as most of the existing applications consider the Transmodel standard, but not the IFOPT extension, which is quite more detailed and also, it is essential for the notion of transfers. Also, we have matched existing data against the resulting metamodel, in order to determine which accessibility elements are included or not. As a result, we have concluded that “pure” IFOPT is not enough to provide all the required information, so it has been necessary to *extend* this model, of course preserving the structure. Therefore, this *extended IFOPT model for accessibility* is another of the contributions of this work.

Afterwards, we have defined an RDF Schema according to both descriptions, and we have used it to capture all the information provided by EMT Madrid in a useful, standard, and linkable form – i.e. as an RDF datastore. We feed this information into our application services, but we also made it available (with necessary restrictions) in several forms, including a Jena-based website to present part of these data in the LOD spirit, and to give a general impression on the way we can exploit this information to provide new applications and services.

8 Future Work

This work has been developed in the context of the CoMobility Project, a general effort in which the analysis and exploitation of the information of transport networks serves to the general purpose of providing a framework to combine public and private transport media, in order to provide both an economic and sustainable way to improve the transit in big cities. This project was already presented in section 2, and much of the future work in this context is related to the integration of the accessibility system presented in the paper with the rest of the subsystems included in this project. Obviously, the general features provided by the whole CoMobility system must have into account if a certain proposed route is accessible or not.

Probably the most important future work in the context of accessibility is to include information about the accessibility of other transport media, such as metro. This way we would be able to check the accessibility of a hybrid route (i.e. including both bus and metro, for instance). This will make possible to check if the extended model we have presented in this work is adequate for several transport media, or if it required further tuning – our current impression is that it is general enough already, but we still have to check that situation. Also, we want to make sure that the existing algorithms apply in this extended scenario – as it should be the case, considering that the information uses a standard format.

Finally, we are also considering the comparison to other formats – in particular, there are very efficient stores which use, e.g. the JSON format. But, at the same time, JSON has not a graph-oriented format, and therefore RDF seems a natural choice in this context. We intend to develop a full comparison and see which one of these systems is actually more efficient. Also, the Jena-based current implementation was intended as an initial prototype. We intend to provide more powerful RDF stores in the near future, though this is mostly an implementation issue.

Acknowledgements

The authors want to thank to Miguel Ángel Garrido Blázquez for his help during the preparation of this article. This work has been supported by the project CoMobility (TIN2012-31104), funded by the Spanish Ministry of Economy and Competitiveness; and it has also been partially supported by the Chair of Ecotransport, Technology and Mobility (http://www.catedraetm.es) at Rey Juan Carlos University.

References

[Becker, 97] Becker, M., Smith, S. F.: An ontology for multi-modal transportation, planning and scheduling. Technical Report CMU-RI-TR-98-15, Robotics Institute, Carnegie Mellon University, Pittsburgh, USA.

[ChETM, 14] Chair of EcoTransport, Technology and Mobility (ChETM) of the Rey Juan Carlos University, accessed Jul 2014: http://www.catedraetm.es/

[Colpaert, 14] Colpaert, P.: Route Planning Using Linked Open Data, The Semantic Web: Trends and Challenges, Lecture Notes in Computer Science Volume 8465, 827-833, 2014.

[CTM, 13] Public Consortium for Transport in Madrid (CTM): Accessibility on Public Transport in Madrid, 2013. Available at: http://www.ctm-madrid.es/

[CTM, 14] Public Consortium for Transport in Madrid (CTM), accessed Oct 2014: http://www.ctm-madrid.es/

[Cuesta, 13] Cuesta, C.E., Cáceres, P., Vela, B., Cavero, J.M.: CoMobility: A Mobile Platform for Transport Sharing, ERCIM News 2013(93), 2013.

[Eclipse, 14] Eclipse, accessed Nov 2014: http://eclipse.org/

[Epsi, 12] e-Public Sector Information Platform (ePSI Platform): Open Transport Data Manifesto, September 2012. Available at: http://www.epsiplatform.eu/

[EMTMadrid, 12] EMT Madrid home page, accessed Sep 2014: http://www.emtmadrid.es

[Fuseki, 14] Fuseki, accessed Oct 2014:
http://jena.apache.org/documentation/serving_data/index.html

[Heath, 11] Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space, Morgan & Claypool, 2011.

[Hobbs, 14] Hobbs, A., Hanley, S.: Big and Open Data in Transport. POST notes POST PN 472, August 2014. Available at: http://www.parliament.uk/briefing-papers/post-pn-472/big-and-open-data-in-transport

[Houda, 10] Houda, M., Khemaja, M., Oliveira, K., Abed, M.: A public transportation ontology to support user travel planning in Proceedings of the Fourth International Conference on Research Challenges in Information Science, 2010.

[IFOPT, 14] IFOPT. Available at: http://www.dft.gov.uk/naptan/ifopt/

[IUEE, 02] Instituto Universitario de Estudios Europeos (IUEE), Universidad Autónoma de Barcelona: La accesibilidad en España. Diagnóstico y bases para un plan integral de supresión de barreras, Col. Estudios e informes. Ministerio de Trabajo y Asuntos Sociales, 2002.

[Jung 12] Jung, J.J.: “Semantic Annotation of Cognitive Map for Knowledge Sharing between Heterogeneous Businesses,” Expert Systems with Applications, 39, 5, 1245-1248, 2012.

[Klyne, 12] Klyne, G., Carroll, J.: Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. Available at: http://www.w3.org/RDF/

[Lakomaa, 13] Lakomaa, E., Kallberg, J.: Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs, IEEE Access 2013 (1), 558-563, 2013.

[LOD, 14] Long, N.H., Jung, J.J.: “Privacy-aware Matching Online Social Identities for Multiple Social Networking Services,” Cybernetics and Systems, 46, 1-2, 69-83, 2015

[Marçal de Oliveira, 13] Marçal de Oliveira, K., Bacha, F., Mnasser, H., Abed, M.: Transportation ontology definition and application for the content personalization of user interfaces, Expert Systems with Applications, 40, 3145–3159, 2013.

[ONCE, 14] Spanish National Society of Blind People (ONCE, Organización Nacional de Ciegos de España), accessed Sep 2014: http://www.once.es/new

[ODUK, 14] London Underground Open Data (ODUK), accessed Nov 2014: http://www.tfl.gov.uk/info-for/open-data-users/

[Pham and Jung, 14] Pham, X.H., Jung, J.J.: “Recommendation System Based on Multilingual Entity Matching on Linked Open Data,” Journal of Intelligent & Fuzzy Systems, 27, 2, 589-599, 2014

[Protegé, 14] Protegé, accessed May 2014: http://protege.stanford.edu/

[RDFS, 14] RDF Schema, accessed Nov 2014: http://www.w3.org/2001/sw/wiki/RDFS

[SparQL, 14] SparQL, accessed Sep 2014: http://www.w3.org/TR/rdf-sparql-query

[Timpf, 02] Timpf, S.: Ontologies of Wayfinding: a traveler's perspective. Networks and Spatial Economics, Kluwer Publishers 2(1): 9-33, 2002.

[Transmodel, 14] European Committee for Standardization: Transmodel, Road Transport and Traffic Telematics. Public Transport. Reference Data Model, EN 12896. Available at: http://www.transmodel.org/index.html

[URI, 14] Wikipedia: URI, accessed May 2014: http://en.wikipedia.org/wiki/Uniform_Resource_Identifier

[Vega, 06] Vega, P.: La accesibilidad del transporte en autobús: Diagnóstico y soluciones, Col. Documentos, Serie Doc. Técnicos, Ministerio de Trabajo y Asuntos Sociales, 2006.