Towards Smart Public Transport Data

A Specific Process to Generate Datasets Containing Public Transport Accessibility Information

Paloma Cáceres, Almudena Sierra-Alonso, Belén Vela, José María Cavero, Carlos E. Cuesta
Escuela Técnica Superior de Ingeniería Informática
Universidad Rey Juan Carlos, URJC
Móstoles (Madrid), Spain
e-mail: {paloma.caceres,almudena.sierra,belen.vela,josemaria.cavero,carlos.cuesta}@urjc.es

***Abstract*—If I was a mother of twins, could I find a route to enable me to get around with my twin baby stroller using public transport? In spite of the relevance of mobility in Smart Cities, there are not many websites or apps with which to generate public transport routes for people with special needs. Of those that do exist, Google Maps is a relevant tool that is employed to calculate routes and find points of interest, while Google Transit Feed Specification (GTFS) is a format used to specify public transport data that allows public transport agents to provide a "feed" complying with this specification. GTFS is employed by Google to calculate a route and represent it on Maps. However, Google Maps does not provide detailed information regarding specific facilities such as accessibility to transit, and GTFS does not specify the structure required to provide that public transport accessibility information. Worse still, it is not easy to find specific and detailed accessibility information that can be downloaded and processed so as to develop services for users. In this work, we propose a systematic method with which to provide accessible public transport datasets in order to improve mobility in Smart Cities. The steps in the process include the extraction of data from the Internet to the generation of open and linked datasets. Moreover, we show a case study in which both datasets for users with special needs and an app prototype that uses those datasets to generate accessible subway routes are developed.**

Keywords-Public transport; accessibility; smart city; open datasets; linked data.

I. INTRODUCTION

Smart cities now appear to be ready to meet the challenge of a more inclusive society that offers the same opportunities for all, and several projects and initiatives have been set up. For example, the Global Initiative for Inclusive Information and Communication Technologies (G3ICT) [1] works to introduce "Smart Cities Programs to Improve Human Rights, Civic Engagement, and Business Outcomes". Another example is World ENABLED (WENABLED) [2], which promotes more inclusive societies working on policies and programs related to human and civil rights. G3ICT and WENABLED are working together on a global strategy for digital inclusion and have identified barriers to smart city accessibility and defined the priority steps required to increase accessibility. One of these steps is "Using open and accessible datasets that include information by and about persons with disabilities" [3]. Smart cities take into account aspects as relevant as, for example education, healthcare, and transportation.

With regard to transportation, we feel that it is necessary to express public transport accessibility and special needs information in a more detailed manner, taking into account that this information is extremely important for a more inclusive society. As we have said, Google Maps is an important tool to calculate routes. To show routes on the map, Google takes data specified in GTFS format. Google Transit Feed Specification (GTFS) [4] defines a "feed": a set of files that describe data concerning public transport. Transport agencies can provide their transport data by means this "feed". Finally, Google employs the "feeds" to calculate the routes and represents them on Maps.

But Google Maps does not provide detailed information about specific facilities such as accessibility to transit, and GTFS does not specify the structure required to provide that public transport accessibility information. In addition, it is not easy to find specific and detailed accessibility information that can be downloaded and processed.

In this work, we propose a systematic method to publish such public transport accessibility data. This method has been validated against real data for the subway in the city of Madrid, Spain (Metro Madrid [5]): we had to validate the generation of accessible subway routes from these data; to do this, we have developed an algorithm which has been implemented as part of the Android prototype app. This prototype (hence the algorithm) makes it possible to obtain an accessible subway route according the Metro Madrid user's needs, a facility that this company does not provide.

The paper is structured as follows: Section 2 presents the state of the art of this work. Section 3 describes our CoMobility and Access@City projects, which are the context of this work. Section 4 introduces some examples of the current state of special needs and accessibility information. Section 5 describes both our proposal, specifying the process used to obtain public transport accessibility datasets, and a case study. Our conclusions and future work are presented in Section 6.

 ISBN: 978-1-61208-622-4

II. Related Works

To the best of our knowledge, currently there are no projects which fully define accessibility limitations and their relationship to the features of public transport.

However, there exist several semantic models which represent specific aspects of the public transport domain. In some cases, they also include accessible wayfinding information or/and accessibility features, or elements for people with special needs or disabilities. Next, we introduce some the more representative among them.

NaPTAN [6] is a vocabulary with which to identify, in a unique manner, the national public transport access nodes of the United Kingdom. It does not incorporate any aspect of accessibility. The ofi-ontology [7], meanwhile, makes it possible to represent whether a place is accessible to people with mobility problems by means of classes such as AccessFacilities and properties such as is_wheelchair_accessible, but it does not take into account other necessary elements in order to provide blind or deaf people with information. DBpedia [8] has a property denominated as isHandicappedAccessible to indicate whether or not a transit station is accessible. Tube Facility ontology [9] incorporates only step free and lift facilities as accessibility elements (mobility disabilities). Accessibility Ontology [10] is similarly also focused on concepts related to supporting only mobility disability problems. TRANSIT [10] is a specific ontology for transit but it does not incorporate any aspects of accessibility to human transport. Landmark Ontology for Hiking [12] is focused on older adults and allows them to walk less through the use of wheelchairs. It formally represents landmarks for hiking. The European OASIS (Open architecture for Accessible Services Integration and Standardisation) project [13] does not incorporate relevant concepts for needs regarding accessibility to public transport.

Transmodel [14] is a European Reference Data Model for Public Transport Information, which provides both a model of public transport concepts and data structures that may be useful when building information systems related to the different kinds of public transport. It does not, however, provide any information about accessibility. IFOPT metamodel [15] was conceived as an extension of Transmodel. It is a prCEN ("Comité Européen de Normalisation –CEN-" in French) Technical Specification and 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 with which to describe accessibility data concerning the equipment of vehicles, stops and access areas.

III. The Context: CoMobility and Access@City Projects

This work is being developed in the context of two related research projects: CoMobility and Access@City.

CoMobility [16] defines a multimodal architecture based on linked open data for sustainable mobility. Its main goals are to improve citizens' mobility, optimize their trips by combining public transport and the sharing of private transport (e.g., car sharing), and also provide a means for accessible trips.

Access@City [17] is a coordinated project that defines a technological framework in which to process, manage and use open data concerning public transport with the goal of promoting its accessibility. One of its subprojects is Multiply@City [18], which is focused on processing and harmonizing public transport accessibility data in a semantic manner, taking into account that data is provided by different sources and will have different formats. It is, therefore, necessary to integrate accessibility data, obtained from open data and by means of Web scraping, and accessible routes, obtained by employing crowdsourcing techniques with those users who use mobile technologies. Figure 1 provides a general depiction of this project.

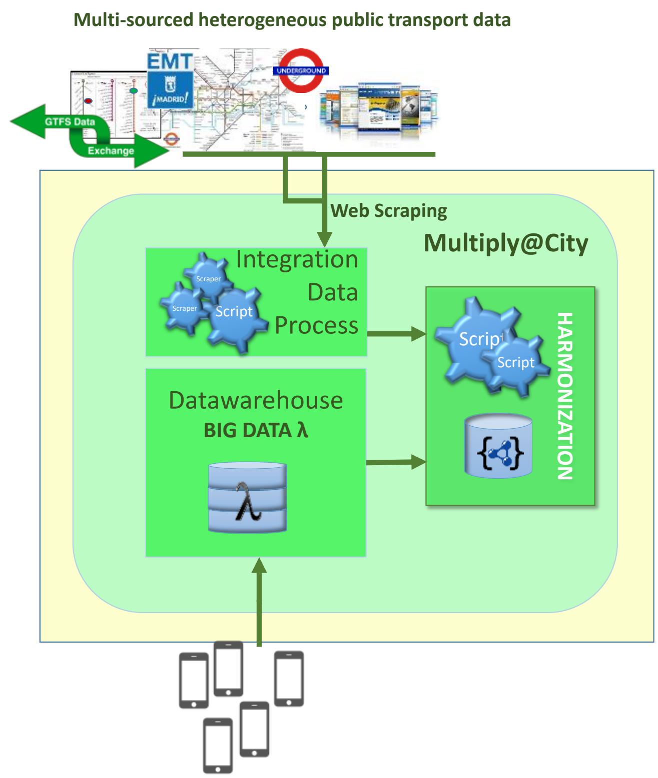

Figure 1 Multiply@City project architecture.

The Regional Consortium for Public Transport in Madrid (CRTM) [19], the Madrid public bus company (EMT Madrid) [20], and the Spanish National Society for the Blind (ONCE) [21] have expressed an interest in the results of our CoMobility and Access@City projects.

IV. THE CURRENT STATE

It is not currently easy for users with special needs to find detailed information regarding public transport on the Internet. In some cases, no accessibility information exists, while in others, it is available only in HTML format (a non-computable format). Only in exceptional cases is it possible to find this information in a computable format.

In this section, we show some examples of how detailed, open and linked the information for users with special needs is. First, we present a study concerning the accessibility information regarding the metro transport media of fourteen European cities. We then go on to describe the GTFS format, which provides the detailed structure employed to specify public transportation schedules and associated geographic information. This can be computable, but has great limitations in terms of specifying transport media accessibility information.

A. Accessibility information of European metro

We have carried out a study on the accessibility of the metro and about how computable that accessibility information is. Our study is based on the information obtained from the official metro websites of several European cities (see TABLE I).

TABLE I SUMMARY OF LINKED OPEN DATA AND ACCESSIBILITY INFORMATION IN EUROPEAN CITIES' METRO TRANSPORT MEDIA

European city	5-star classification	Accessibility information
Barcelona	***	×
Berlin	*	×
Hamburg	*	√
Istanbul	*	√
London	**	√
Madrid	*	√
Milan	*	√
Moscow	***	×
Munich	*	√
Oslo	*	×
Paris	*	×
Rome	*	√
Saint Petersburg	*	√
Vienna	*	×

We have focused on the fourteen cities, which have the largest metro networks -in kilometers-. These cities are, in alphabetical order: Barcelona, Berlin, Hamburg, Istanbul, London, Madrid, Milan, Moscow, Munich, Oslo, Paris, Rome, Saint Petersburg and Vienna. We have then analyzed how open, linked and computable the data on the websites are, that is, how many stars they have attained in the *linked open data 5-star* classification of Tim Berners-Lee [22]. To finish, we have checked, for each city, whether accessibility data exists on the website. We concluded that:

a) no city has four or five stars for the data shown on its website, that is, there is no linked open data on the metro companies' official websites;

b) 10/14 cities provide open data of the lowest category (the one-star category);

c) 3/14 cities only provide data as regards the three star category;

d) 8/14 cities show transport media accessibility.

It is, therefore, possible to conclude that it is not easy to find relevant information about public transport accessibility on the Internet.

B. Accessibility information of GTFS

GTFS [4] is a format for public transportation schedules and associated geographic information. A feed of GTFS is a collection of a maximum of six CSV files, with a .txt extension. Only two of these files currently include some sort of information regarding accessibility and special needs: *stops.txt* and *trips.txt*.

With regard to the *stops.txt* file, GTFS states that "A stop is a location where vehicles stop to pick up or drop off passengers. Stops can be grouped together, such as when there are multiple stops within a single station. This is done by defining one stop for the station, and defining it as a parent for all the stops it contains. Stops may also have zone identifiers, to group them together into zones". The *stops.txt* file includes an optional column denominated as `wheelchair_boarding` to indicate accessibility to this kind of information about a stop. GTFS states that "It identifies whether wheelchair boardings are possible from the specified stop or station."

With regard to the *trips.txt* file, GTFS states that "A trip represents a journey taken by a vehicle through stops. So, a single trip represents one journey along a transit line or route". This file includes two columns, both optional, related to accessibility limitations or special needs: `wheelchair_accessible` and `bikes_allowed`. In this work, we address only accessibility aspects and will not, therefore, discuss the `bikes_allowed` field.

In both files, the accessibility information provided by GTFS refers only to mobility impairment, while other impairment characteristics have not been considered: the special needs of deaf and blind people. In a previous work, we proposed to enhance accessibility information in Google Maps by adding new pieces of information to GTFS [23].

V. OUR PROPOSAL

The intent of our proposal is to improve accessibility information in order to support new social accessibility services, such as calculating public transport routes that are accessible to all. The eventual objective is to provide datasets as open data so as to support these services.

We work with the Identification of Fixed Objects in Public Transport (IFOPT [15]) reference datamodel. It is a

 ISBN: 978-1-61208-622-4

prCEN ("Comité Européen de Normalisation –CEN-" in French) Technical Specification and, as mentioned previously, 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 with which to describe accessibility data concerning the equipment of vehicles, stops and access areas.

In this work, we describe the process defined in order to generate a public transport dataset from public transport data and to make it accessible. This process consists of the following steps:

1. Study the accessibility features of the public transport network included on the official website of the transport media.
2. Identify the data semantics of this accessibility information in order to align them with a reference vocabulary (MAnto [24]);
3. If the alignment is possible, obtain the original data and semantically annotate them; if the alignment is not possible, analyze the difference between the original data and the reference vocabulary in order to extend it and, in this case, return to the second step;
4. Use the dataset as a service to improve the lives of citizens and public transport users.

In the following subsection, we validate the process against real data for the subway in the city of Madrid, Spain (Metro Madrid [2]).

A. A case study: Collecting and annotating data concerning Metro Madrid in Spain

In order to carry out the case study, it was necessary to use real data concerning Metro Madrid. We applied the four steps defined in the aforementioned process:

- First, we studied the lines, the stations on each line and the accessibility features. Metro Madrid specifies accessibility in two different ways. The first is on the *stations* webpage, where the user can determine whether a station is accessible and the kind of accessibility (universal or partial). The second is on the specific *accessible stations* webpage, where the user can determine the kind of accessibility of each station (universal or partial). More information on this website can be found in [5].
- Second, we identified the data semantics of the Metro Madrid accessibility features so as to align them with MAnto. TABLE II shows the correspondence between both.

TABLE II RELATIONSHIP BETWEEN METRO MADRID ACCESSIBILITY AND MANTO.

Accessibility stations and quays - Metro Madrid-	Accessibility Limitations -MAnto-
Universal accessibility	WheelchairAccess *true* AudibleSignsAvailable *true* StepFreeAccess *unknown* LiftFreeAccess *false* EscalatorFreeAccess *unknown* VisualSignsAvailable *true*
Complementary accessibility measures without lifts and/or ramps	WheelchairAccess *false* AudibleSignsAvailable *true* StepFreeAccess *unknown* LiftFreeAccess *true* EscalatorFreeAccess *unknown* VisualSignsAvailable *true*
Lifts and/or ramps without complementary accessibility measures	WheelchairAccess *true* AudibleSignsAvailable *false* StepFreeAccess *unknown* LiftFreeAccess *false* EscalatorFreeAccess *unknown* VisualSignsAvailable *false*

- Third, as Metro Madrid does not provide the means required to download these data, we developed a scraper to determine which stations are accessible and their kind of accessibility in each case. This information was semantically annotated and stored in RDF/XML (Resource Description Framework/eXtensible Markup Language) [25]. The accessibility information concerning Metro Madrid is specified at quay level. Each stop place has two quays. For example, at the Sol stop place, only quay 2 on line 1 is not accessible for wheelchairs, as is shown in Figure 2.

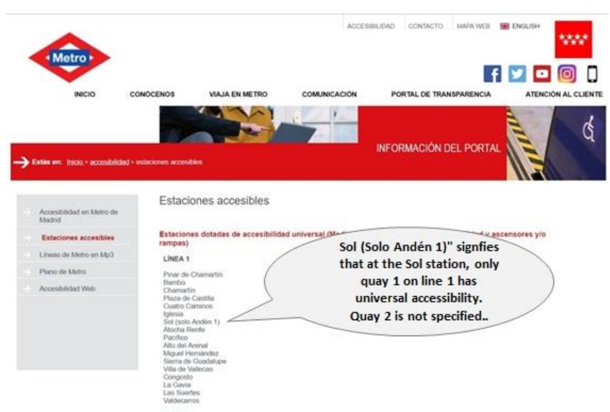

Figure 2 Quay 2 on line 1 is not accessible for wheelchairs at the Sol stop place.

The following piece of the code describes the accessibility for Metro Madrid, focusing on the Sol stop place and the two quays on lines number 1 and 2 (see Figure 3).

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-
  ns#>.
@prefix mao: <com.vortic3.MANTO#>.

mao:StopPlace_Sol rdf:type mao:StopPlace ;
 mao:StopPlaceName "Sol" ;
 mao:StopPlace_Quay_has mao:Quay011 ;
 mao:StopPlace_Quay_has mao:Quay012 ;
 mao:StopPlace_Quay_has mao:Quay021 ;
 mao:StopPlace_Quay_has mao:Quay022 .

mao:Quay011 rdf:type mao:Quay ;
 mao:QuayName "Line1Quay1" ;
 mao:AccessibilityLimitation_WheelchairAccess "true";
 mao:AccessibilityLimitation_AudibleSignsAvailable
  "true";
 mao:AccessibilityLimitation_StepFreeAccess "unknown";
 mao:AccessibilityLimitation_LiftFreeAccess "false";
 mao:AccessibilityLimitation_EscalatorFreeAccess
  "unknown";
 mao:AccessibilityLimitation_VisualSignsAvailable "true".

mao:Quay012 rdf:type mao:Quay ;
 mao:QuayName "Line1Quay2" ;
 mao:AccessibilityLimitation_WheelchairAccess "false";
 mao:AccessibilityLimitation_AudibleSignsAvailable
  "true";
 mao:AccessibilityLimitation_StepFreeAccess "unknown";
 mao:AccessibilityLimitation_LiftFreeAccess "false";
 mao:AccessibilityLimitation_EscalatorFreeAccess
  "unknown";
 mao:AccessibilityLimitation_VisualSignsAvailable "true".

mao:Quay021 rdf:type mao:Quay ;
 mao:QuayName "Line2Quay1" ;
 mao:AccessibilityLimitation_WheelchairAccess "true" ;
 mao:AccessibilityLimitation_AudibleSignsAvailable"true";
 mao:AccessibilityLimitation_StepFreeAccess "unknown";
 mao:AccessibilityLimitation_LiftFreeAccess "false";
 mao:AccessibilityLimitation_EscalatorFreeAccess
  "unknown";
 mao:AccessibilityLimitation_VisualSignsAvailable "true".

mao:Quay022  rdf:type mao:Quay ;
 mao:QuayName "Line2Quay2" ;
 mao:AccessibilityLimitation_WheelchairAccess "true" ;
 mao:AccessibilityLimitation_AudibleSignsAvailable
  "true";
 mao:AccessibilityLimitation_StepFreeAccess "unknown";
 mao:AccessibilityLimitation_LiftFreeAccess "false";
 mao:AccessibilityLimitation_EscalatorFreeAccess
  "unknown";
 mao:AccessibilityLimitation_VisualSignsAvailable "true".
```

Figure 3 Accessibility of Metro Madrid by means of MAnto.

- Fourth, the dataset was used in the Android prototype app. It makes possible for users to obtain an accessible route on the metro by means of a specific algorithm, taking into account their different needs. Perhaps no accessible route exists, but if one does, then this app will be able to calculate it. The user interface in Figure 4a shows the information the prototype requires in order to establish the route: first, the origin and the destination of the route; second, the list of user's special needs (people with special mobility needs, blind people, hearing-impaired people), and third, the characteristics of the route (minimizing commutes or stations). The list of special needs is related to the columns in TABLE II. As previously mentioned, the first column identifies universal accessibility (everyone can travel), no lifts or ramps (blind and hearing-impaired people can travel, but mobility impaired people cannot), and only lifts and ramps (people with special mobility needs can travel, but blind or deaf people cannot).

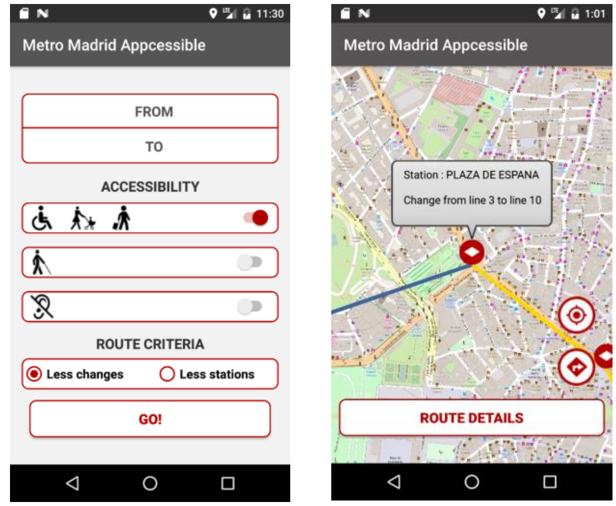

Figure 4 (a) App user interface requesting user needs; (b) Calculated accessible route.

Such accessibility features must match the special needs as shown in the prototype app. An example of an accessible route (for people with special mobility needs) is also shown in Figure 4b.

We have to underline that the app is just a prototype. A prototype because it does not include: (1) an accessible user interface; (2) downloading live updates or incident locations in the Metro Madrid network; (3) uploading updates or incidents identified by the user of the app.

The purpose of this prototype is to generate subway routes from the Metro Madrid data according to the specific user's needs; in this way, it includes: (a) a user interface in which a user can select his/her user's needs (only those that the Metro Madrid data can support); (b) the generation of accessible subway routes from the origin to a destination according to the user's needs.

With regard to the validation of the app, we have tested that the generated routes are effectively correct (verified in real-life), that is, that the routes can be used by people with specific disabilities identified in the user's needs.

VI. Conclusion and Future Work

One of the major goals of smart cities is to achieve an inclusive society through the use of technologies. In this context, one important challenge is to facilitate the mobility of all citizens. If this challenge is to be met, information regarding means of transport and their accessibility features is, therefore, required in order to be able to provide accessible routes for every user, including special needs users.

As shown in this paper, these accessibility data are not easy to find and cannot be processed owing to the diversity of sources and formats. We have presented a study of the state of accessibility data in several European cities and

 ISBN: 978-1-61208-622-4

shown that GTFS does not provide the constructions required to represent accessibility data. In order to resolve this issue, we have proposed a systematic process with which to gather the information, unify the formats and semantically annotate it: the first step is carried out by means of a scrapper; in the second and the third steps, we represent the data in RDF so as to unify and annotate the data for their further use in applications that help citizens move in a city.

As future work, we intend to improve the systematic process defined in this work in order to provide fully open datasets of different public transport networks. Moreover, these datasets must be published in an open platform that will permit free access to accessibility and special needs data. The process will take into account the accessible routes generated using crowdsourcing techniques. These routes will be generated by users with special needs. We have to developed different apps for these kinds of users with the aim of capturing the geographical information and the accessibility characteristics of these routes. This information will be incorporated as smart data into the Multiply@City platform (see Figure 1). Moreover, the apps still have to be validated and will be developed following the existing standards and initiatives of accessible user interfaces. The same directives have to be applied to the protoype app presented in this paper.

ACKNOWLEDGMENT

This work is supported by the Multiply@City project (TIN2016-78103-C2-1-R) and the Access@City project (TIN2016-78103-C2-1-R), funded by the Spanish Ministry of Economy and Competitiveness.

We would like to thank Miguel Maroto for his support.

REFERENCES

[1] The Global Initiatives for ICTs (G3ict), http://g3ict.org/resource_center/g3ict_smart_cities_initiative, [retrieved: March, 2018].

[2] World ENABLED, http://worldenabled.org, [retrieved: March, 2018].

[3] Smart Cities for All, a Global Strategy for Digital Inclusion proposed by G3ict and World ENABLED, March 2017.

[4] Google Transit Feed Specification, https://developers.google.com/transit/gtfs/, [retrieved: March, 2018].

[5] Metro MADRID, http://www.metromadrid.es/en/, [retrieved: March, 2018].

[6] NAPTAN, http://transport.data.gov.uk/def/naptan [retrieved: March, 2018].

[7] Ofi-ontology, http://ip-kom.net/data/html_documentation_ontologien/ofi-ontology.html [retrieved: March, 2018].

[8] DBpedia,http://mappings.dbpedia.org/index.php/OntologyProperty:IsHandicappedAccessible, [retrieved: March, 2018].

[9] TubeFacility, https://github.com/AccessAtECS/RailGB/blob/master/standard/public/ajax/tube.php [retrieved: March, 2018].

[10] Accessibility ontology, https://github.com/ox-it/accessibility-ontology/blob/master/accessibility-ontology.ttl [retrieved: March, 2018].

[11] TRANSIT, http://vocab.org/transit, [retrieved: March, 2018].

[12] T. Sarjakoski, P. Kettunen,H. Halkosaari, M. Laakso, M. Rönneberg, H. Stigmar, Landmarks and a Hiking Ontology to Support Wayfinding in a National Park During Different Seasons. In M. Raubal, D. M. Mark, and A. U. Frank (Eds.), Cognitive and Linguistic Aspects of Geographic Space, Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. DOI:10.1007/978-3-642-34359-9.

[13] OASIS, http://www.oasis-project.eu/ [retrieved: March, 2018].

[14] Transmodel, standardization, E. C. (2016). Transmodel, Road Transport and Traffic Telematics. Public Transport. Ref. Data Model, EN 12896, http://www.transmodel.org/en/cadre1.html, [retrieved: March, 2018].

[15] CEN/TC 278. Intelligent transport systems - Public transport - Identification of Fixed Objects In Public Transport (IFOPT), EN 28701:2012.

[16] C. E. Cuesta, P. Cáceres, B. Vela, and J. M. Cavero, "CoMobility: A Mobile Platform for Transport Sharing", ERCIM News 2013(93), pp. 22-23, 2013.

[17] Access@City project (TIN2016-78103-C2-1-R), funded by the Spanish Ministry of Economy and Competitiveness.

[18] Multiply@City project (TIN2016-78103-C2-1-R), funded by the Spanish Ministry of Economy and Competitiveness.

[19] Regional Consortium for Public Transports of Madrid (CRTM), http://datos.crtm.es/, [retrieved: March, 2018].

[20] EMT Madrid home page, http://www.emtmadrid.es/Index.aspx?lang=en-GB, [retrieved: March, 2018].

[21] Spanish National Society of Blind People (ONCE, Organización Nacional de Ciegos de España) http://www.once.es/new/otras-webs/english, [retrieved: March, 2018].

[22] 5-star classification, Tim Berners-Lee, http://5stardata.info/en/, [retrieved: March, 2018].

[23] P. Cáceres, A. Sierra-Alonso, C. E. Cuesta, J. M. Cavero, and B. Vela, "Enhancing Accessibility Information in Google Maps: Adding new pieces of information to GTFS to improve accessibility" In Proceedings on the Second International Conference on Universal Accessibility in the Internet of Things and Smart Environments (SMART ACCESSIBILITY), pp. 6-9, Nice (France), 2017.

[24] Mobility and Accessibility Ontology (MAnto): https://vortic3.github.io/MANTO/, [retrieved: March, 2018].

[25] RDF/XML (Resource Description Framework/eXtensible Markup Language), https://www.w3.org/TR/rdf-syntax-grammar/, [retrieved: March, 2018].