Received September 21, 2020, accepted October 7, 2020, date of publication October 26, 2020, date of current version November 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3033584

Improving Urban Mobility by Defining a Smart Data Integration Platform

PALOMA CÁCERES, ALMUDENA SIERRA-ALONSO, CARLOS E. CUESTA, BELEN VELA, AND JOSÉ MARÍA CAVERO BARCA
Computer Science and Statistics Department, Universidad Rey Juan Carlos, 28933 Móstoles, Spain

Corresponding author: Paloma Cáceres (paloma.caceres@urjc.es)

This work was supported in part by the Access@City research project, funded by the Spanish Ministry of Science, Innovation and Universities under Grant TIN2016-78103-C2-1-R.

ABSTRACT One of the key factors employed to define the well-being of citizens in the urban environment is mobility, since it defines a set of flows and connections that constrain those citizens' individual and collective behaviour. However, the complexity of this activity on the scale of a city makes this a complex problem in computational terms. One of the main reasons for this is the asymmetry of information: different actors have access only to partial or outdated information, and many relevant data are simply unavailable. In this article, we propose a *data integration* architecture and platform with which to combine relevant data from many different sources and provide the results in a variety of forms. This integration uses semantic technologies, thus ensuring that the relationships among data show their actual meaning and are appropriately interpreted. The resulting platform amalgamates: open data, which is available from public sources; extracted data, obtained from public sites by means of scraping techniques; pre-processed data, stored in public databases; aggregated data, acquired from pervasive devices by means of crowdsourcing; smart data, supplied by mobile applications and enriched with contextual information, or data concerning specific incidents, often provided by the users themselves. The semantic integration of this information makes it possible to compute a wide range of results, from accessible transport routes to identifiable events, in a coordinated manner. The general public is then supplied with these results through the use of specific software, via either mobile applications or the web. We are of the opinion that the collective use of this information may improve urban welfare.

INDEX TERMS Crowdsourcing, data acquisition, data processing, open data, pervasive computing, semantics, smart data, social computing, software architecture, urban mobility.

I. INTRODUCTION

The concept of the *smart city* implies the notion of a better city: one in which data is gathered in order to learn about the problems in that city and decide upon potential solutions to them; one in which there are fewer disruptions, the citizens' lives are improved, and their experience is enhanced – in summary, a city whose inhabitants' well-being increases. The urban landscape is a highly complex ecosystem and can easily degenerate into a hostile environment; as this complexity grows, the general public is beginning to rely on *technology* to make it evolve in the right direction. This is, allegedly, the actual purpose of smart cities: using technology to solve the challenges of the evolving urban space, and applying software solutions in order to tackle both large-scale issues and small-scale concerns.

One of the key factors is urban mobility: the way in which people move through the city defines their perception of the environment, the concentration of the population and the interferences and interactions among individuals; in short, it is the framework that delimits their behaviour. Urban mobility is delimited by the combination of pedestrian routes, private transport and public transport. The first is mostly relevant over short distances, while the second cannot be controlled but only restricted. Public transport is, however, the backbone that defines the flows of people and goods within a living city. An appropriate regulation of public transport has direct consequences as regards improving citizens' welfare – and in this case, the notion of *welfare* must be understood in the most

The associate editor coordinating the review of this manuscript and approving it for publication was Miltiadis Lytras.

general sense, not limited to (but including) the economic perspective.

The use of public transport, as opposed to private transport, provides great advantages in terms of the entire urban community's well-being. On the one hand, traffic jams, energy consumption and carbon emissions are reduced. On the other, new parking infrastructures are not necessary, which allows the construction of new pedestrian spaces, such as parks and recreation areas, in their place. Any means employed to increase the use of public transport and improve its efficiency have a positive effect on the whole system.

Much of the current research on smart cities has been related to cyber-physical systems and the Internet of Things, and much of it has, therefore, had to deal with hardware issues. However, the *smartness* of the city is, in reality, provided by the software that processes the data captured by that hardware. Indeed, on the scale of a city, the actual issue is that of capturing all kinds of information in different data streams, synchronizing them, processing them, and performing the corresponding analyses in order to learn about the current situation of the city, even in real time, and decide which actions would serve to improve that situation.

In this context, rather than isolated efforts to process separate fragments, an integral approach is relevant and required. The data context of the city is defined by both *stable* information and a *dynamic* status. The first of these provides *structure*, which can be recovered from conventional databases and includes data concerning the city itself, the means of transport and the combined routes they define. The second comprises *variable* data streams obtained from different sources, and includes information regarding unexpected events, temporary barriers, accidents and incidents, or simply the flow of individuals within the city at a particular moment: collapsed streets, traffic jams or public demonstrations. Many proposals concentrate their efforts on finding routes within the stable structure, without considering the critical influence of temporary obstacles; and of course, the influence of a certain event cannot be estimated without considering the remaining data. Both aspects of this information are, therefore, relevant, and both must be considered in a coordinated manner, even when considering *synchronization*; e.g. an accident may cause a traffic jam, but several hours later, this is no longer relevant; a multitude might cause a standstill, but an alternative route could provide a detour around it. When dealing with a dynamic status, the *timing* is as relevant as the place itself.

Moreover, there is an enormous variety of information sources. Apart from the abovementioned emphasis on sensor-oriented computation and IoT devices, there are many other options. First, it is necessary to consider that any smart city initiative also places emphasis on transparency and on the publishing of *open data* concerning the city itself. These open data sources (many of which employ a form of processable Linked Open Data) may vary in their approach, ranging from an automatic data stream, perhaps originating from a physical sensor, to more elaborate and even pre-processed data sources, such as the periodic emission of structured data files concerning, e.g. the pollution levels in the city. Second, many other sources, which provide information in the form of an accessible API (application programme interface), are not exactly considered part of an open data initiative but share much of its philosophy. Third, the largest potential source of information is probably not even these initiatives, but rather the presence of computational devices throughout the city, that is, the citizens' *smartphones* (and other smart devices) with their many sensors and their huge computational capabilities. If adequate software is employed, these smartphones can be the means to both capture and even pre-process information that is not accessible in any other form, and provide the smart city's inhabitants with the processed information in a useful form. This implication of the citizens themselves in the building of the *smart* city not only defines a way in which to extend the "sensor network" in that city, but also creates a *crowdsourcing* endeavour, which may take many different forms, and in which each new software application in fact embodies a different conception. It could even be argued that this crowdsourcing approach is required for the success of any smart city initiative, and it should be considered as an essential part of any complex *urban computing* approach.

Fourth, and finally, there are the considerable number large databases that store all the *stable* information mentioned above, and this information is the basic mainstay of the city's information processing system.

Coordinating all of these sources of information is highly complex, and in many situations goes beyond what a computational system can deal with and enters the scope of the so-called "big data" initiatives. Nevertheless, the *integration* of even a small part of these data may have a huge impact on the well-being of the individuals affected. Even partial improvements can have very significant consequences.

It is for this reason that our work proposes:

- A data integration architecture and platform, on which all the data from the different sources are stored, *semantically* annotated and harmonised, thus ensuring a feature-rich integration; and
- A set of specific and focused software applications designed for both smartphones and the web that can solve specific issues (accessibility, incidents and infrastructure) one at a time.

We intend to provide a generic architecture so as to support data integration on an urban computing platform. This will lay the foundations for many different applications, each of which will be designed in order to *improve the well-being* of urban citizens and to *enrich* the platform itself, thereby allowing more sophisticated decisions to be made. The key aspect as regards providing this generic integration is the use of semantic technologies.

The main contributions of our work in addition to the *Services on the move for* Urban mobility (SofUR) platform are, therefore, the following:

- A *Services on the Move* Architecture (SoMA), which is structured in three following different subsystems:

- An *Open Data & Acquisition subsystem*, which obtains data from the open and public sources.
- A *Semantic subsystem*, which processes and integrates data and then transforms them into semantic data by using our domain-specific MAnto ontology.
- A *Smart Data subsystem*, which gathers smart data concerning public transport and its accessibility by using various apps developed for this purpose.

- A suite of applications that offer urban services to citizens with the aim of putting mobility in the city *on the move*. This suite is divided into the two following groups
  - The *suite of mobile applications*, which is a set of mobility applications designed for mobile devices.
  - The *suite of Web applications*, which supplies information services to data consumers, who will use those data according to their needs or convenience, and which also offers mobility services on the Internet.

This article is structured as follows. Section 2 shows a review and discussion of some similar proposals, while Section 3 provides a summary of the different technologies used throughout this work and related efforts. A detailed description of the structure of our proposal (the Services *on the move* for Urbanmobility –SofUR- platform and the underlying SoMaarchitecture) is then shown in Section 4, which also includes an inclusive enumeration of the associated suite of applications. Section 5 illustrates how these applications prove the validity of the proposa, in addition to providing our conclusions and an initial account of some future work.

II. RELATED WORK

In this section, we discuss some of the most representative software approaches employed to improve citizens' well-being as regards urban mobility. Firstly, we examine how, within the framework of smart cities, urban mobility is considered to be one of the factors affecting that well-being.

The term smart city has been used for over two decades [3]. Numerous definitions have been proposed during this period; however, while they have commonalities, they also differ in some respects. Choubari *et al.* [27] and Albino *et al.* [3] present a compilation of these definitions.

Some authors consider that the defining characteristic of a smart city is the use of ICTs when applied to citizens' daily lives [43], [44], [82]. This implies that it is necessary to collect data by means of sensors, meters, appliances, personal devices, or similar, in order to later integrate them and, through their use, provide citizens with services and assist them to make informed decisions. But a different trend has recently appeared, which claims that other relevant aspects are required in order to consider a city as a smart city. Of the definitions mentioned above [3], [27], some include not only ICTs, but also aspects such as *sustainability* and *liveability*. There is currently a tendency to consider that a multidisciplinary approach is required [53] in order to make the transition to a smart city.

In this respect, several authors have drawn up lists of dimensions in order to determine which aspects actually constitute a smart city [3], [27], [40], [54], [60]. Although each list contains different dimensions, there are several in common, such as people (human), social aspects, the economy, mobility, quality of life or the natural environment and technology.

Within the framework of smart cities and in line with the objective of improving the quality of citizens' lives, it is worth emphasizing the need to improve urban mobility, as mentioned in the introduction. In their definition of a smart city, Giffinger *et al.* [40] mention mobility as a basic element to be taken into account during its construction. Washburn *et al.* [82] also confirm the need to make services such as transport more efficient, and this is directly related to citizens' mobility. Albino *et al.* [3] state that "high-quality and more efficient public transport that responds to economic needs and connects labour with employment is considered a key element for city growth". Lombardi *et al.* [51] also opine that smart mobility is a basic component of a smart city.

Indeed, urban mobility has, in recent years, become an object of study in an attempt to improve it in order to increase citizens' well-being [77]. One of the main aspects of this well-being is public transport. The European Union has funded numerous projects in order to address this issue [1], [35]. These projects develop infrastructures with which to improve public transport and make it more comfortable and more accessible to all, and they incorporate the use of information technologies (IT) to make public transport more efficient and effective and to ensure that its use is a more pleasant experience.

In this last respect, several software applications have been developed that inform users about the public transport network and how to use it. These applications provide information about the lines, stops, accesses, etc. of the different means of public transport. Almost all of them provide the user with the possibility of requesting a route (optimal in terms of time or distance) from an origin point to a destination point. In some cases, they include information about accessibility features, or elements for people with special needs or disabilities. When considering the importance that people and social aspects acquire in the general context and conception of smart cities, it should not be forgotten that an essential feature must be the inclusion of all social groups by overcoming barriers of language, culture, education and disabilities [3], [53]. Taking accessibility into account is, therefore, essential if people with special needs are to be provided with a good experience as regards the use of public transport. Several pieces of software use crowdsourcing to gather data about public transport. The information collected is then made available to other users in order to improve the responses that citizens attain from these applications.

Those applications that take aspects of accessibility into account usually focus on a particular type of impairment or group of people. Wheelmate [84] or WheelMap [83] consequently provide information regarding wheelchair accessibility for many places in 45 countries. In addition, users can modify the degree of accessibility of registered places or mark new ones. Cardonha *et al.* [23] have developed an application that enables users to incorporate accessibility information so as to collaborate in the creation of accessibility maps. Access Map [1] helps people with mobility needs to plan an accessible route in the city of Seattle. These applications consider only mobility limitations. Others, such as Landmark Ontology for Hiking [71], are intended for elderly people. In this case, the application aids them to walk less when hiking. There are also applications that consider other types of special needs, one of which is "*CiudadesPatrimonio de la Humanidad*" [15], a web application that provides accessibility information concerning tourist routes in World Heritage cities. This information refers to mobility needs, in addition to others related to vision and hearing. These applications are not, however, generally customisable: their users can establish the place to which they wish to go, but cannot establish what their accessibility needs are.

As stated above, the European Union is making a great effort to improve public transport [35], and it is necessary to highlight two projects that focus on the use of IT to improve mobility on urban public transport: ACCESS 2 ALL [2] and Mediate [55]. ACCESS 2 ALL exhaustively analyses users' possible needs with respect to public transport. By taking these needs into account, it establishes guidelines to ensure that urban public transport is accessible to all citizens. Moreover, it proposes customised services for route guidance. The Mediate Project [52] has identified a set of measures with which to describe the degree of accessibility of a particular means of transport and has developed an application with which to quantify it. Another result of Mediate Project has been a Good Practice Guide for accessibility. These two projects seek to establish a theoretical framework that will cover all aspects of citizens' mobility, but do not provide solutions in the form of user applications.

Another aspect that qualifies smart mobility is the use of data [47]. In order to achieve actual smart mobility, services have to rely on a lot of information: not only a certain amount of data, but also a significant amount of variety. For effective mobility, these data must be up-to-date, and in many cases, real-time information is required. A paradigmatic example is the communication of incidents in the public transport network. If an application calculates an accessibility route, but, for example, the recommended lift does not work, then the route is not useful for the user. In order to obtain this information in real time, it is possible to take advantage of the widespread use of smartphones and encourage the users themselves to provide pieces of this incident information when they encounter them. This is referred to as crowdsourcing. There are also applications that use crowdsourcing to improve urban public transport. For example, Tiramisu Transit [85] tells the user whether there is room left for a wheelchair on a bus or how full that bus is. Moovit [59], meanwhile, provides information about the status of each service, calculates routes and indicates when to get off. The OneBusAway [33] project consists of a set of tools whose intention is, among other goals, to comply with the bus schedules or to decrease waiting times in order to increase well-being on urban public transport. This objective can be extended to other transit systems. In addition, OneBusAway permits users to make comments about these tools. Swiftly [75] provides more accurate vehicle arrival data for transport agencies, thus enabling them to provide their users with better information and allow them to better plan their journeys.

Other initiatives that use crowdsourcing are the BUSUP project [14] and the CIVITAS initiative [24]. The former allows users to book crowdsourced buses on demand, while CIVITAS is an initiative from the EU to promote a new urban mobility culture. One of its mobility strategies includes safe and secure transport for all users, taking into account a variety of needs [35]. In the CIVITAS initiative, several pilot projects are being deployed in European cities to test new accessibility concepts, such as smart access facilities for wheelchairs [35].

OpenTripPlanner (OTP) [63] is a project that calculates routes combining different transit systems, including bicycles and walking routes, and takes (transport) accessibility into account. OTP obtains data from OpenStreetMap [64] and GTFS (General Transit Feed Specification) feeds [41].

As will be noted, there are numerous approaches with which to improve users' well-being when they use urban public transport. However, to the best of our knowledge, none of them considers the calculation of routes by taking the accessibility features for any type of need into account or maintains real time data concerning temporary incidents (a lift does not work, there is work taking place that hinders blind people, etc.) in the public transport network.

It is now necessary to discuss the need for data with which to provide intelligent urban public transport. These data originate from heterogeneous sources, including data scattered on the Web in several formats or data originating from portable devices (e.g., smartphones) [50]. In order to obtain greater benefit from existing data, they should be published in a coordinated manner, harmonised and linked, thus uniting the efforts of public and private initiatives. The management, maintenance and publication of data have, therefore, become a growing challenge. Several initiatives have, however, been set up to meet this challenge, some of which consist of governmental data portals based on CKAN [25]. CKAN is a data management system that facilitates the publishing and sharing of data. Other works present platforms on which to manage, link or publish data from different sources, e.g., QuerioCity [52], AECIS [36], the proposal by Bischof *et al.* [12] or ATIS [7]. Some proposals develop more comprehensive frameworks covering several stages of the data publication process, such as Santos *et al.*'s process [70], the VIVO approach [81], the linked data platform LinkedLab [30], the CITIESData framework [50] or CityPulse [65]. All of them handle data from different domains, but none of them focuses on public transport data and less still on the accessibility of public transport. Several works can also be found in the domain of urban public transport. The proposal by Schlingensiepen *et al.* consists of designing a framework for autonomic transport in the smart city environment, but it is for transport in general [72]. Ning *et al.* present a control system for urban rail based on artificial systems, considering human factors in order to improve the experience during the use of urban rail, but it is not intended to allow users to personalise the use of the data [61]. However, Lau and Ismail's framework employs a crowdsourcing approach to provide passengers with real time data in order to meet their needs [49]. But once again, nothing about accessibility is mentioned.

We can conclude that there are many works in the area of smart technologies whose objective is to improve urban mobility so as to increase the well-being of citizens, but that there is still much to be done, especially as regards accessibility, if urban mobility is to be accessible to all.

Table 1 summarises the related work. This table presents the issues analysed in this work in order to improve urban mobility and, therefore, the well-being of the citizens in the context of intelligent cities.

TABLE 1. Related Work Classified by Issues Related to Urban Mobility in Smart Cities.

<table>
<tr><td>Definition and characterization of smart cities</td><td colspan="2">Albino et al. [3]
Choubari et al. [27]
Hall [43]
Harrison et al. [44]
Lytras and Visvizi[53]
Washburn et al. [82]</td></tr>
<tr><td>Definition of dimensions in order to characterise a smart city</td><td colspan="2">Albino et al. [3]
Choubari et al. [27]
Giffinger et al. [40]
Lytras, Visvizi and Sarirete[54][53]
Nam and Pardo [60]</td></tr>
<tr><td>Urban mobility as a factor in the well-being of the citizens in a smart city</td><td colspan="2">Albino et al. [3]
Giffinger et al. [40]
Lombardi et al. [51]
Washburn et al. [82]
Tomaszewska and Florea[77]</td></tr>
<tr><td>The role of public transport in urban mobility</td><td colspan="2">Accessmap. [Online]. Available: https://www.accessmap.io.[1]
Gaggi,Fluhrer and T. Janitzek [35]</td></tr>
<tr><td>ICT applications with which to improve the experience of using public transport</td><td colspan="2">Accessmap. [Online]. Available: https://www.accessmap.io.[1]
Access2All [2]
Ciudades Patrimonio [15]
Cardonha et al. [23]
Mediate [55]
OpenTripPlanner[63]
Sarjakoski et al. [71]
Wheelmap[83]
Wheelmate[84]</td></tr>
<tr><td rowspan="2">The need for data in order to provide intelligent urban mobility</td><td>Collection (crowdsourcing)</td><td>BUSUP [14]
CIVITAS [24]
Ferris et al. [33]
Lau and Ismail [49]
Swiftly [75]
Zimmermanet al. [85]</td></tr>
<tr><td>Publication and harmonization</td><td>Bischof et al. [12]
CKAN [25]
Darari et al. [30]
Gandon et al.[36]
Liu et al. [50]
López et al. [52]
Puiu et al. [65]
Santos et al. [70]
VIVO [81]</td></tr>
</table>

III. RELATED TECHNOLOGIES

The work described herein was carried out using various technologies. The most important technologies used to gather and harmonise data and to develop the server are described below.

A. WEB SCRAPING

Data are scattered on the Web, distributed on many different sites. Furthermore, they have different formats that are structured to a greater or lesser extent. Data can appear as plain text or organised in html tables. It is possible to download files containing data in XML or JSON format or to invoke APIs that also provide files containing data in those or other formats. For these data to be useful, they have to be collected, related to each other and given a common format.

Web scraping solves that need by obtaining data from the Internet. It consists of gathering data automatically (from the Internet) by using programmes (denominated as bots or scrapers) that query a web server, request data, parse those data to extract the information required and store them for their later use [57].

There is a variety of programming techniques, programming languages and libraries with which to attain the data distributed on the web. One of these languages is Python [65]. Several modules can be used to programme a scraper. Three of the most frequently used are [42]:

- The **requests** library [69], which is a wrapper over the **urllib** Python library whose goal is to handle the HTTP requests. Its use is recommended by the Python 3 documentation, since it is friendlier than **urllib** [67].
- The **Beautiful Soup** library [9] helps to extract the information of interest from HTML or XML files. It can be used with different parsers depending on the format to be analysed. This library provides a function that creates a tree from the HTML or XML file. Each node in this tree contains a tag. These tags, which contain the information of interest, are searched by means of other functions, also provided by the library. Once the tag is located in the tree, its contents are extracted.
- **Scrapy** [73] is a more powerful framework than Beautiful Soup and can be scaled, but is more difficult to use. Scrapy [48] is an open source framework with which to extract data from websites using XPath 26]. It offers the tools for the efficient extraction of unstructured data that

are scattered throughout the web, their processing and their storage in the structure and format required [48].

Scraped data are integrated using several formats, mainly XML, JSON or CSV.

B. SEMANTIC TECHNOLOGIES

In this subsection, we introduce the various semantic technologies that are necessary to understand our work.

The Resource Description Framework (RDF) is a model for data interchange on the Web [68]. In the RDF, data are described as a set of triples. A triple is formed of three components: a subject, a predicate and an object. The subject is a resource that is related to an object by means of a predicate. The object can be another resource or a literal. The predicate can describe a relationship between the subject and object or a property of the subject [29].

In order for the data of which the triples are formed to be referenced, the RDF identifies each component of the triple with a URI (Uniform Resource Identifier) [11]. The set of triples can be represented graphically as a graph (an RDF Graph) on which subjects and objects are nodes and predicates are represented with edges. For all the users of that data to be able to interpret the predicate in the same way, the use of standard vocabularies is recommended, since they will allow the definition of the predicates of the triples. Finally, the RDF graph is serialised to enable computers to handle data. One of the most frequently used notations is RDF-XML [36], which is based on an XML format, while SPARQL is used to recover data from an RDF model [62].

SPARQL is an RDF graph query language that has been standardised by the World Wide Web Consortium (W3C) [62]. The use of SPARQL allows the definition of queries, which search for data in the RDF graph that satisfy the conditions of that query. It is a key technology in the development of the semantic web, since the data structured in RDF format could not otherwise be accessed. A SPARQL query recoveries RDF sentences, that is, triples. As with SQL, it is necessary to distinguish between the query language and the engine for data storage and retrieval. There are consequently several implementations of SPARQL that are generally linked to different technologies.

One of these implementations is Apache FUSEKI [34], which forms part of the Apache Jena project. It is a SPARQL server and provides the means to recover and update data from a JENA repository using SPARQL.

Apache JENA [4] is a free and open source Java framework that is employed to build applications and which accesses the Semantic Web and Linked Data. JENA is used to define an RDFS [13] so as to describe the underlying relationships that exist between data on the RDF graph. JENA provides a set of functions and services in which to store, extract and publish data as RDF triples that comply with the RDF schema.

RML (the RDF Mapping Language) is a general language that permits the definition of rules with which to map heterogeneous data sources onto RDF graphs [31]. Mapping rules are defined in order to transform data from, for example, XML, JSON and CSV formats into the RDF. The rules expressed by means of RML transform heterogeneous data structures into an RDF data model. These rules transform the data format by defining new triples, semantically annotated with the vocabularies (ontologies) specific to the data domain, and integrate those triples into RDF graphs already defined in the same domain. The rules can be defined from any data source, in formats as diverse as CSV, HTML, XML, or even a format similar to that of relational databases. The result of applying the rules is a set of RDF triplets that uses the predicates and vocabulary types of the domain.

C. DATA PROCESSING TECHNOLOGIES

In this subsection, we introduce the various technologies related to our proposal.

Apache Kafka [5] is a stream-processing software platform. It implements a ‘‘publish/subscribe’’ message system, is scalable and distributed and provides a valuable platform on which to process streaming data between applications.

Apache NiFi [6] is an easy-to-use, powerful and reliable system with which to process and distribute data. It makes it possible to automate the movement of data between different systems quickly, easily and securely. It can load data from different sources, and has a very powerful web interface that allows its users to visually design the data flow, act on the process and monitor that process.

MongoDB [58] is a scalable, flexible and distributed document database. It stores data in JSON-like documents, thus making data integration easier and faster. Searches can be carried out by means of fields, range queries, regular expressions or JavaScript functions. It supports field indexing and real time aggregation. MongoDB provides great availability in the form of replica sets. It can run on multiple servers, and balances the load or duplicates data in order to keep the system up and running in the case of failure.

D. WEB INTEGRATION TECHNOLOGY

In this subsection, we introduce the various technologies that are necessary for web integration.

Spring Boot [74] is a framework that is used to develop applications in Java. It provides a set of tools in order to configure and programme applications on any platform. The use of Spring Boot allows the programmer to focus only on the solution and forget about the complex tasks of configuration in terms of data access, security, communications, etc.

Apache Tomcat [7] is open source software that permits web applications to be deployed in an integrated manner by sharing the same subsystem. Some of our applications are directly run on top of Tomcat and do not, in these cases, employ Spring.

IV. SERVICES ON THE MOVE FOR URBAN MOBILITY

As stated in the related work section, several solutions regarding how to deal with these issues already exist but, to the best our knowledge, there is no comprehensive solution that solves most of the known problems from a single perspective.

Our *Services on the move for* Urban mobility (SofUR) platform defines a set of services whose objective is to improve the mobility of all the citizens in a city by promoting the use of public transport. This results in an improved quality of life and welfare in cities with regard to air quality, the flow of traffic, etc.

We have, therefore, developed a service-based architecture and a suite of software applications, which together comprise the aforementioned platform. The *Services on the Move* Architecture (SoMA) is structured in three different subsystems that carry out acquisition, processing and integrating activities with data obtained from crowdsourcing, public and opensources and that provide enriched information in the form of semantic data. The suite of software applications offers different urban mobility information services obtained from these semantic data for both citizens and data consumers and, in some cases, also provides smart data by means of crowdsourcing techniques.

In the following subsections, we first describe the *Services on the Move* architecture, after which we specify the different subsystems in which our proposal is structured. Finally, we present the suite of applications developed that offers a set of services to citizens and data consumers in order to improve public transport users' well-being.

A. SERVICES ON THE MOVE ARCHITECTURE

As mentioned above, this work is being developed with the aim of providing public transport and its corresponding accessibility information. We, therefore, acquire data concerning public transport from different open and public sources, after which we process the data obtained. We then integrate and enrich them, and finally, we transform these data into semantic data.

In order to carry out these activities, we have designed an extensible structure, the Services *on the Move* Architecture (SoMA), which is composed of three subsystems. The first, denominated as the *Open Data & Acquisition subsystem*, is responsible for obtaining data from the open and public sources, while the second, the *Semantic subsystem*, processes and integrates data and then transforms them into semantic data by using our domain-specific MAnto ontology. The third, denominated as the *Smart Data subsystem*, gathers smart data concerning public transport and its accessibility by using various apps developed for this purpose, and those data are then sent to the Semantic subsystem in order to transform them into semantic data.

Fig. 1 shows the general structure of the SofUR proposal and, within it, our SoMAarchitecture, in which the above-mentioned subsystems are represented. Each of these subsystems will be explained in greater detail in subsections B, C and D. The figure also includes the suite of apps (mobility and web apps), which will be explained in subsection E, and which are an integral part of the SofUR platform.

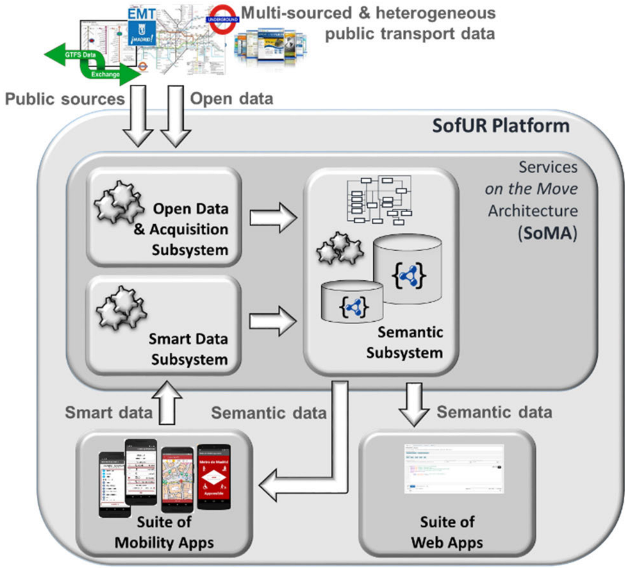


FIGURE 1. SofUR platform.

B. OPEN DATA & ACQUISITION SUBSYSTEM

The Open Data & Acquisition (ODA) subsystem is responsible for processing information obtained from public sources and open data related to the public transport infrastructure and its corresponding accessibility features. Fig. 2 shows the details of this subsystem.

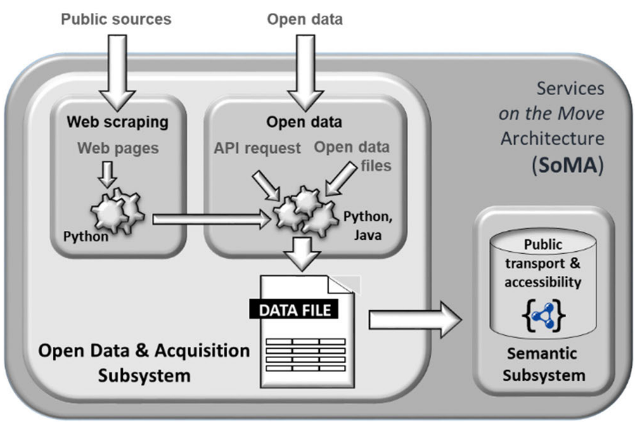


FIGURE 2. Open data & acquisition subsystem.

The first kinds of sources are *public information sources*, in which interesting public transport data are usually scattered on the Web and distributed on many different sites. All these data are collected through the use of web scraping techniques and we have, therefore, programmed several scrapers using the Python language. The Scrapy [73] and Beautiful Soup [9] libraries, mentioned above, are used to access the websites, and to seek and extract data of interest. We focus on the data extraction and processing of the existing information on the web concerning public transport, along with its accessibility, by means of a method for the semi-automatic generation of a data scraper for the public transport domain. This method allows the extraction of public transport data and the existing accessibility information from a selected website. Moreover, we have developed a web tool that applies the aforementioned method in order to generate a data scraper for the public transport domain [80]. The data collected are organised and related according to an underlying domain model, in this case a public transport domain model, which was described in detail in [80]. Finally, we convert them into a structured data file with a common format (usually JSON or CSV) and these files are transferred to the Semantic Subsystem.

The second kinds of sources are open data sources, which provide open access to public transport data, either through an API service, or by directly downloading data files. Each source provides data in a different format (CSV, XML, KML) and with diverse internal structures, but all of them must satisfy the constraints of a domain model. In many cases, sources offer open transport data as CSV files by following GTFS. Those open data files can be processed generically, because they have the same structure as that shown in [21]. With regard to the sources without a GTFS structure, it is necessary to process each one of them in a specific way in order to seek the information of interest, which is specified in the domain model, and taking the particular structure of each source into account. We process sources of both types by means of Java and Python programmes in order to extract the specific data required for our purpose. The main reason for using Java was initially that we had more experience with this language, and also owing to its expressive power. We have since used Python more frequently, essentially owing to its readability, simplicity [45] and the availability of libraries. Finally, we convert them into a structured data file with a common format (usually JSON and/or CSV, depending on the origin) and send it to the Semantic Subsystem.

C. SEMANTIC SUBSYSTEM

The Semantic (SEM) subsystem is responsible for processing structured data files from the ODA and Smart Data (SMD) subsystems, converting them into RDF graphs by means of programmatic (Java) processing or, more recently, RML transformations, and then storing them in the Jena repository. The Jena repository is a single repository, but it stores two different data collections: the *public transport & accessibility* collection and the *incidents* collection. Details of the SEM subsystem are provided in Fig. 3.

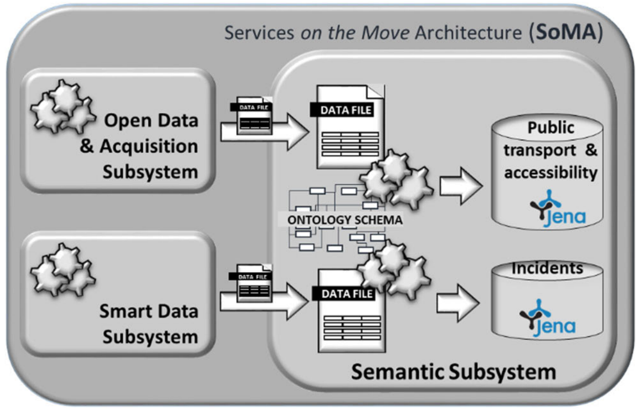


FIGURE 3. Semantic subsystem.

The SEM subsystem applies an ontological schema, denominated as the MAnto schema, which is related to the public transport infrastructure and the accessibility feature domain. The definition of the MAnto schema is based on the Transmodel (European Reference Data Model for Public Transport Information) [76] and IFOPT (Identification of Fixed Objects in Public Transport) [45] reference data models. Transmodel describes a model of both public transport concepts and data structures related to the different kinds of public transport. IFOPT extends it by including specific structures designed to specify accessibility data concerning the equipment of vehicles, stops and access areas in order to define a model for the principal so-called *fixed objects*, i.e. elements related to the access to public transport (e.g., entrances, stop places, connection links). We have, therefore, obtained a set of metadata (MAntoterminology) with which to semantically annotate the public transport infrastructure and its accessibility elements data from sources [21].

As mentioned in the previous subsection, data from the different sources are organised and related according to an underlying conceptual model, which we have defined as the domain model. Our data are, therefore, always domain objects and are related in the same way, that is, by following the relationships described in that model. The ontological schema, denominated as the MAnto *schema*, is also based on the same domain model. It represents data in that context, i.e. in that domain. The MAnto *vocabulary* (a set of terms) consequently also represents the domain objects, and the relationships between terms are additionally contemplated in that reference model, again our domain model. Fig.4 shows how the MAnto ontology schema is defined following the structure of the domain model in the transport context.

Fig. 4 shows only a fragment of the domain model in (A) and a fragment of the MAnto ontology schema in (B). We depict how an object in the domain model, denominated as *oD-n*, is defined as a class in the MAnto ontology schema. For example, the Public Transport object in (A) is defined as the http://com.vortic3.MANTO#PublicTransport class in (B). Similarly, a relationship between two objects in the domain model, denominated as rD-x, is defined as a data property in the MAnto ontology schema. For example, in (A), the relationship between the Public Transport and the Line object makes it possible to establish which lines belonging to a means of Public Transport (Metro Madrid, the London Underground, etc.) are defined as the http://com.vortic3.MANTO#TransportFor data property in (B). It also specifies that this property domain is the http://com.vortic3.MANTO#Line class, and its range is the http://com.vortic3.MANTO#PublicTransport class.

As defined in [80], all data sources in the public transport context are specified according to the same underlying

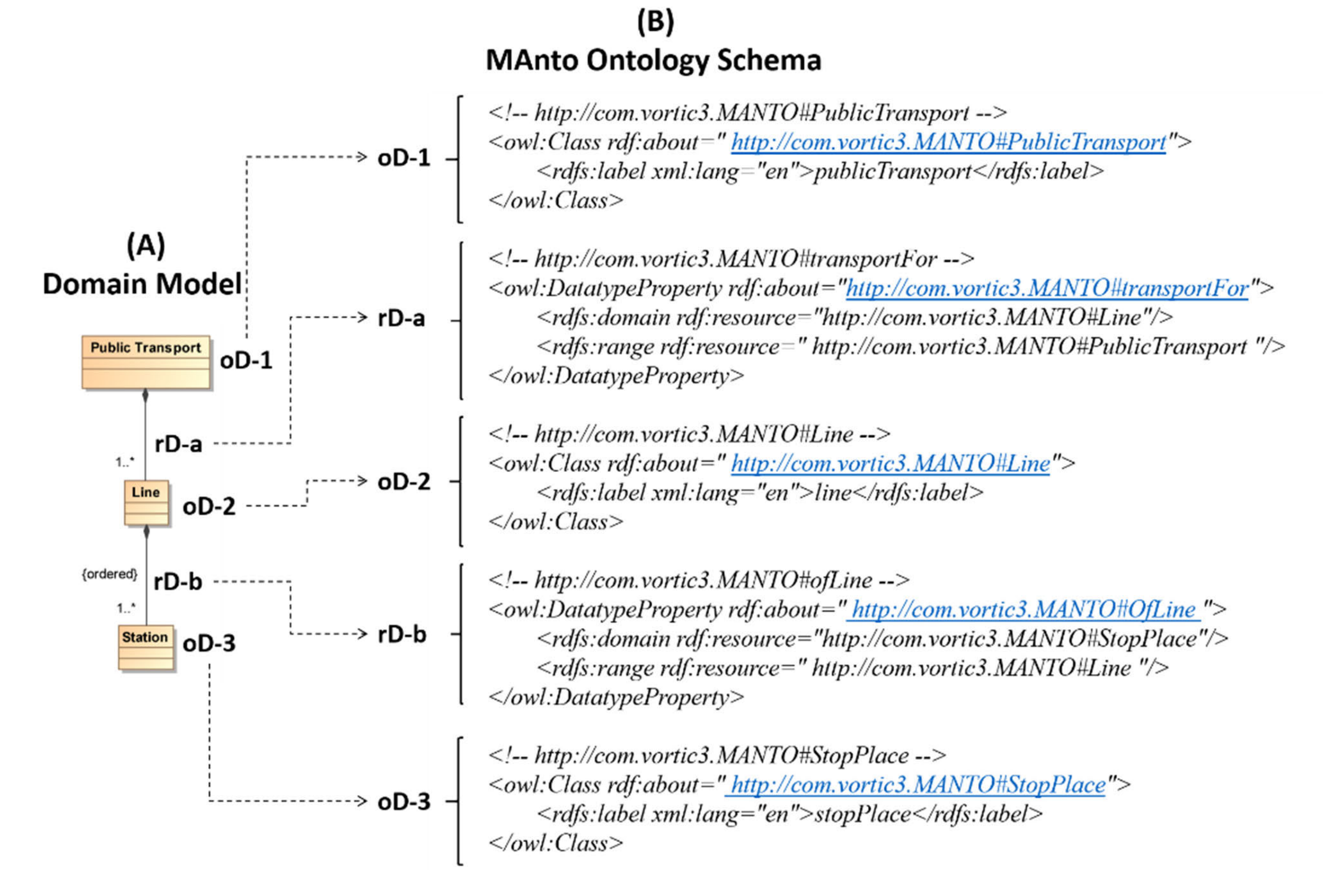


FIGURE 4. The definition of MAnto ontology schema.

domain model. As our approach uses the same domain model to identify the objects from data sources and to define the MAnto ontology schema, it is easy to deduce which terms from the MAnto ontology are mapped onto the existing objects in the data sources.

In fact, in the SEM subsystem, we work with the data file generated from the OAD subsystem. Once again, the data from that file also represent the same objects from the domain model. Moreover, we can use the MAnto ontology to define a set of consistent mappings with which to *semantically annotate* the data from the structured data file generated by the OAD subsystem. An example of a semantically annotated data fragment is shown in Fig.5.

In this figure, it will be noted that the subway has a line (*sch:name "Line 1"*) that is internally coded as line 1_4, and which is composed of a set of stations (par_4_295, par_4_294, par_4_29) ordered as a sequence.

At the end of the process, these data will be stored in the semantic Jena repository, in the *public transport & accessibility* collection, as shown in Fig.3. Data from this collection can be downloaded at any moment from http://coruscant.my.to:8080/download/metro.xml.

```
mao:line 1_4  a  mao:line ;
    mao:sequence [ a <#Seq> ;
          <#_1> mao:par_4_295 ;
          <#_2> mao:par_4_294 ;
          <#_3> mao:par_4_293 ;
          <#_4>mao:par_4_27 ;
          .......
                            ] ;
          mao:transportFor "Metro" ;
          sch:description "Pinar de Chamartín-Valdecaros" ;
          sch:name "Line 1" .
```

FIGURE 5. Data annotated by means of MAnto ontology.

The available data is now growing steadily, as this information is increasingly more valuable to users. In this respect, it might be necessary to change the ontologies in order to adapt to this growth if the domain model evolves. In fact, MAnto is a 'living ontology', that is, it can be expanded through the use of an iterative and incremental process [39], thus making it possible to semantically annotate new data that could appear over time.

The SEM subsystem similarly includes other ontological schema related to incidents (i.e. changes in the working state) of the accessibility elements in the public transport network, which could occur at any time (more information is provided in [16]). Incident data, which are obtained from the Smart Data subsystem, have to be semantically annotated and then stored in the Jena repository, in the *incidents* collection, as shown in Fig. 3. In this case, we have defined a set of mapping rules with which to automatically transform data from the SMD subsystem into RDF graphs by means of RML. Sample data from the *incidents* collection can be downloaded from http://coruscant.my.to:8080/download/events.xml.

D. SMART DATA SUBSYSTEM

The Smart Data (SMD) subsystem employs crowdsourcing techniques [21] to process real-time data from different applications that we have developed for this purpose. This real-time information is related to (a) *incidents*, that is to say, variations in the working state of the accessibility elements of the public transport that may occur at any time, and (b) *pedestrian routes* performed and requested by users themselves.

One example of data in this context is a lift that does not work at a specific transport station, which is part of a specific transport line. This event also has an associated opening date. When the incident is solved, we can close it by indicating the final date. The event data concerning accessibility features, which are processed from many sources (i.e. a crowd), are stored in a data file (see Fig. 4). The information in this data file is relevant in order to update the current accessibility features in the network. This data is our SMD output.

The SMD subsystem then provides the SEM subsystem with the processed data. Details of the SMD subsystem are shown in Fig. 6.

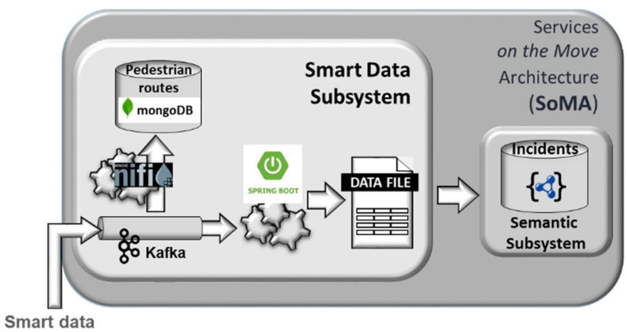


FIGURE 6. **Smart data subsystem.**

We use Apache Kafka technology to manage and gather smart data streams, that it to say, flows of incidents regarding the working state of public transport accessibility features (which usually arrive in emergent surges, and could became massive in size) and specific pedestrian routes to be followed by users. Incidents are then processed by a specific Java transformation, which is supported by a Spring Boot substrate, in order to generate the corresponding structured data and send them to the SEM subsystem. Pedestrian routes are, meanwhile, processed by means of a NiFi configuration, and are stored in MongoDB to be quickly reused; this information has not yet been semantically annotated.

E. THE SUITE OF APPLICATIONS

We have developed a suite of applications, that is to say, several software applications, that offer urban services to citizens with the aim of putting mobility into the city *on the move*. The suite is divided into two different groups, the first of which is the *suite of mobile applications* and is, as its name suggests, a set of mobility applications designed for mobile devices. Some of these devices provide the SMD subsystem with smart data, while others use the MongoDB support. The second group is a *suite of Web applications*, which not only offers mobility services on the Internet, but also supplies information services to data consumers, who will use those data according to their needs or convenience.

A detailed description of these application suites is provided below.

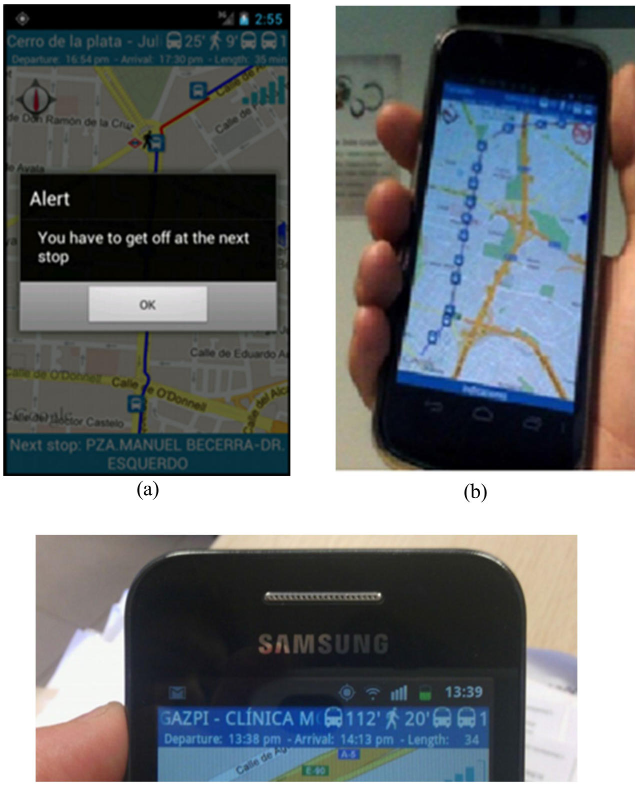
1) SUITE OF MOBILE APPLICATIONS

The suite of mobile applications is a set of five different Android apps. Three of them provide bus and metro mobility services for the city of Madrid in Spain. These services compute specific routes on which some of their users' special needs will be taken into account. One of these three apps also allows users to indicate incidents related to the accessibility elements in the public transport network, that it to say, to indicate their working state. The aim of the fourth app, which is based on gamification techniques, is to capture the accessibility elements in the public transport network and its working state, among other functionalities. Finally, the fifth app registers the user's pedestrian route, on which his/her special needs can be indicated. We should stress that the last three apps also collect data using crowdsourcing techniques, with the users' permission. A detailed description of each of them is provided as follows.

Notify.me [38] provides users of the EMT Madrid company with two different urban bus route services [32]: guiding the user along the bus route and notifying the users about the events on the route. EMT Madrid is the urban bus public company of the city of Madrid and provides an open data platform via web services [28]. The app accesses them in order to obtain the specific urban bus route requested by the user. It then shows it on a map, after which it guides the user along the route and notifies her when a relevant stop is reached by means of acoustic, visual and sensorial messages. Fig. 7a shows an example of a visual message. Fig. 7b displays a route on the map, on which each bus stop is identified by means of a blue icon. Fig. 7c shows the corresponding information for this route at the top of the user interface. The first line shows the starting point (LEGAZPI), the destination (CLÍNICA MONCLOA), the time by bus (112 minutes), the time on foot (20 minutes) and a single transfer. The second line shows the departure and arrival time (11:38 and 14:13) how long this route is (34 km).

We wish to highlight that this software is currently employed by EMT Madrid buses Company [79].

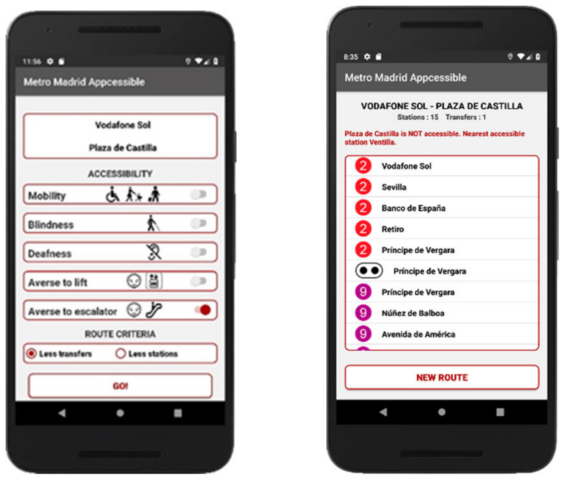
Another application is *MMAppcessible*, which offers Metro de Madrid (MM) users accessible routes on this means



(a) (b) (c)

FIGURE 7. Notify.me user interface.

FIGURE 8. MMAppcessible user interface.



(a) (b)

FIGURE 9. MMA4A screen.

of transport. MM is the public metro company of the city of Madrid (Spain).

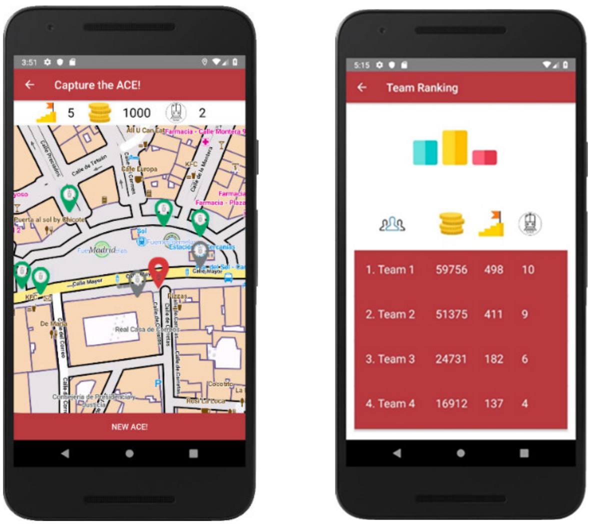
This app, therefore, works with the locally stored *public transport & accessibility* collection, which was previously downloaded from the semantic repository. *MMAppcessible* permits users to indicate the origin and destination of their route, any special needs (visual, auditory or mobility limitations –for example, pushing a pram, or using crutches or a wheelchair-) and their preferences, such as fewer transfers or stops. *MMAppcessible* then shows the route, if it exists, on a map and permits the users to navigate along it. Fig. 8 shows the user interface.

This app won an award in the open programming competition at the School of Computer Science and Engineering at Rey Juan Carlos University, in Spain.

MMA4A [16] is an improved version of MMAppcessible that takes all the user's special needs into account thanks to the IFOPT specification regarding this means of transport. Fig. 9 shows two app screens. That in Fig. 9a is the screen used to request a route and on which it is possible to indicate the user's need (in this case "averse to escalators"). That in Fig. 9b, meanwhile, shows the route and, in this case, the fact that the destination "Plaza de Castilla" is not accessible for the user, who has an aversion to escalators. The recommended destination route is, therefore, Ventilla.

When the app starts up, it requests information from the server regarding whether the *public transport & accessibility* and the *incidents* collections are up to date with respect to the locally stored collections. With regard to the route calculation, the *MMA4A* now works with the two collections, because the incidents on the MM infrastructure could allow it to modify the route as regards the user's needs. As mentioned above, the accessibility elements are part of the infrastructure of MM, and the route (with its corresponding accessibility elements) must be adapted to the user's needs. For example, a person with a temporary mobility issue needs a lift to gain access to the public transport network. It is, therefore, necessary to know which accessibility elements are operational at each station. It is for this reason that the notification of incidents merits special attention: when a user detects an incident, s/he can notify it by means of the app and *MMA4A*, therefore, requests the accessibility element that does not work (lift, escalator, steps, ramp or travellator) and the station at which it is located. The app then sends the corresponding incident to the server to be included in the *incidents* collection of the semantic repository. Other users will be able to benefit from this information, which may be vital if they are to follow a route according to their special needs.

Access 'n' Go! [17] is an Android application that has been developed using gamification and crowdsourcing techniques, and permits its users (the players) to capture public transport accessibility elements (ACEs) such as lifts, escalators, etc. The app then provides the SofUR proposal with smart data concerning the working state of ACEs, that is to say, updated information in real-time regarding those elements. Fig. 10a shows the main user interface of *Access 'n' Go!*, on which the red icon is the player's current position, a green icon represents an ACE that could be captured and a grey icon represents one that has already been captured. The game level, points and badges achieved by the player are shown at the top of the screen. Fig. 10b shows the team ranking user interface.



(a) (b)

FIGURE 10. Access 'n' Go! screen.

The objective of *AccessPedRo* [78] is to obtain the pedestrian route that the user should follow. The users must, therefore, first register and give their consent to this activity (see Fig. 11a). The app then requests the users' needs for the route (see Fig. 11b). From that moment, *AccessPedRo* obtains the users' locations on the routes they are following. The app eventually permits the users to visualise their routes on a map (see Fig. 11c).

The smart data generated by the app is then sent to the SMD subsystem. Although *AccessPedRo* is still in the testing phase, it can already be used to compute accessible pedestrian routes in its current form.

2) SUITE OF WEB APPLICATIONS

We have developed two different Web applications, which have different purposes and audiences.

The first Web app, denominated as MMAR, provides citizens, including those with special needs, with public transport infrastructure and accessibility information services

(a) (b) (c)

FIGURE 11. AccessPedRo screens.

concerning the Metro de Madrid (MM) company in order to allow them to get around using this means of transport. MM is the public metro company of the city of Madrid in Spain [56]. This Web app permits its users to know what specific accessibility features each specific MM station has. Moreover, it provides the possibility of calculating routes based on the users' special needs and the accessibility features of the MM network. The Web app is available at http://bit.ly/MMAR_1. Fig. 12 shows its user interface.

The second app, denominated as MMSemQuery, is an open access Fuseki endpoint, in which the dataset can be queried by means of SPARQL queries. MMSemQuery will normally be used by data consumers, who can download specific MM data in different formats from the dataset. This app is available at http://coruscant.my.to:3030/dataset.html. The user interface of MMSemQuery is shown in Fig. 13.

V. CHALLENGES ADDRESSED

The development of this data integration platform has involved addressing different challenges and overcoming several difficulties. The greatest effort is related to the integration and harmonisation of data, since the origin, format and completeness of the data is highly diverse. In addition, in order to carry out this harmonisation, it has been necessary to previously define an ontological scheme based on transport domain standards which, as mentioned above in section 4.C, are IFOPT [46] and Transmodel [76]. The ontological scheme

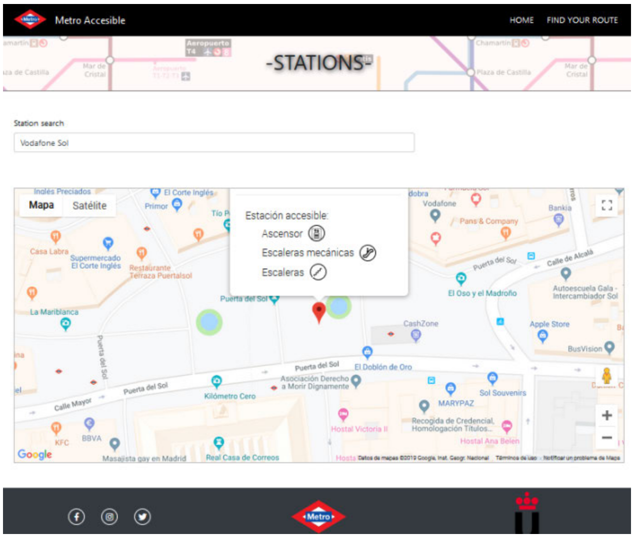


FIGURE 12. MMARuser interface.

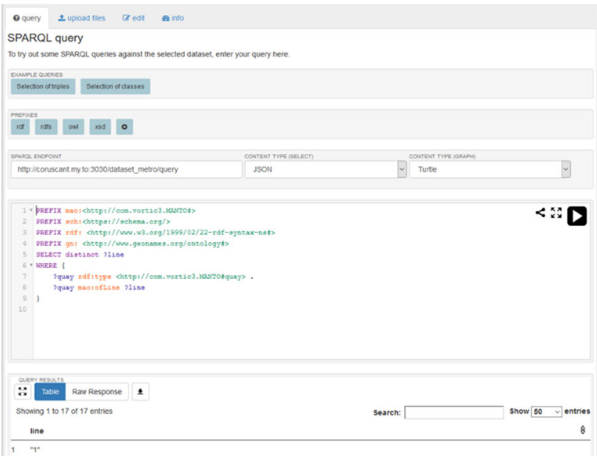


FIGURE 13. MMSemQueryuser interface.

has been defined by respecting the structure, syntax and semantics of the standards considered.

The most relevant aspects of this effort, which have been divided into two categories ((a) format, content and completeness data, and (b) semantic and granularity) are discussed below.

With regard to the first category:

- Different transport media have several types of infrastructures (stops, stations, entrances) and logistics (vehicles and routes). For some means of transport, such as the Metro, this infrastructure is part of their responsibility, and for others, such as the bus service, the infrastructure is the responsibility of other institutions, such as the city council. Accessibility features in these media are, therefore, provided by the institutions themselves, by other institutions or are, in other cases, non-existent. The transport data for a specific means of transport may, therefore, originate from different sources, and the transport information offered is consequently *not uniform* in terms of format, content and completeness.
- The same means of transport are managed by different companies in different cities, signifying that the terminology and characteristics of their infrastructures and logistics are different in each case. For example, a Metro Madrid line has only one single route, while a London Tube line can have more than one route. The transport information provided is, again, not uniform in terms of format, content and completeness.
- Some of the transport companies provide their data in an open format, such as the London Tube. In this case, the information is more complete and structured than that from companies that do not provide it, such as Metro Madrid; this signifies that it must be obtained by means of scraping techniques. Moreover, this information sometimes has to be completed with additional data from other sources; it would, for example, be necessary to add the geographical coordinates to the entrances and stop places of Metro Madrid.

With regard to the second category:

- Each transport company offers the information with different levels of granularity. For example, one company may indicate that a station has universal accessibility, while another may also indicate its specific accessibility infrastructure (lift, ramp, escalator, stairs, etc.).
- The meanings of the terminology used may differ. For example, some transport companies state that they provide universal accessibility by noting that they serve people with visual, hearing and/or mobility problems. However, IFOPT considers that phobias to certain infrastructures are also a factor to take into account when referring to universal accessibility. For example, a person with a phobia to lifts needs to have access to stairs, and the availability of a lift does not, therefore, represent universal accessibility for that individual.

The ontological scheme has been designed by taking into account both the reference standards (Transmodel, IFOPT) and all of the aforementioned problems in order to harmonise the great variety of data managed in terms of formats, contents and meanings. This ontological scheme is constantly evolving, since all of the above-mentioned cases must, according to our analysis of data from other transport companies, be considered.

As will be observed, the applications developed have addressed these problems. For example, if we consider different means of transport, such as *Notify.me* for an urban bus service (Fig. 7) and *MMAppcessible* for a metro (Fig. 8), different meanings associated with universal accessibility have been included, e.g. the inclusion of passenger phobias, as in the case of MMA4A (Fig. 9), etc. In summary, all the challenges and issues mentioned in this section have been considered during the conception and development of this architecture and the associated platform; in some cases, several times.

VI. VALIDATION

Our proposal is basically a *data integration architecture* (the *Services on the Move Architecture*, SoMA); as described previously, this architecture gathers data from many different and diverse sources: linked *open* data, public APIs, processable *open* data files, data scraped from dissimilar websites, data collected from *smart* devices, data retrieved from public databases, *spatial* data available on public maps, topographical data recovered from non-open public sources, *smart* data pre-processed by mobile phones, data streams provided by IoT devices, *open* data streams, or pre-processed data aggregations of previous crowdsourcing efforts. Joining all these data together, reconciling their divergences and complementarities, and semantically harmonising their contents is both an enormous and a rather *abstract* effort. It is simultaneously very specific, as all the features are derived from real-world constraints. The magnitude of the effort is related to different aspects of the elements involved: means of transport have diverse infrastructures, which are mostly permanent (stops, stations, entrances), and several logistics (vehicles and routes), which are usually much more variable; and both categories have accessibility features, which are provided on an individual basis – many instances of the same element may have different accessibility issues. Furthermore, accessibility has a ''weakest-link'' approach: an accessible element (e.g. a train) which passes through a non-accessible location (e.g. a station) is, in fact, no longer accessible. It is particularly difficult to validate this work, as this implies both evaluating its suitability and showing its benefits.

As discussed in Section 1, the benefits of the architecture are directly related to its usefulness and its direct impact on the well-being of citizens, and obviously on the impact on the collective use of public transport and its consequences for the dynamics of the *smart city* itself. Many of the inconveniences that appear in the urban mobility system are owing to asymmetries in the access to relevant information; the availability of a unified integrated *data hub* gathering much of this relevant information has a direct influence on improving the everyday life of those involved. A perfect example of this could be our previous efforts [19] focusing on *accessibility data* in public transport, and the tangible usefulness of these results in the related groups of interest, such as citizens with reduced mobility, or impaired people.

As mentioned in the Introduction, it is unfeasible to attempt to solve all the issues of public transport simultaneously; however, each separate effort in this direction provides an immediate benefit, not only for direct participants, but also for the collective common welfare. Our approach as regards both the building of the architecture and its validation is, therefore, to consider each particular problem and solve it in isolation. This is, however, done by *using all the existing information* in the solution and *integrating any acquired data* into the repository, or any single application into the platform. This means that each of the specific applications described in Section 4 is an integral part of the global system, and thus also serves to validate its overall utility. By validating every single local application, we, therefore, also validate the full global system. Indeed, each of these specific applications (the suite of mobile and web applications) uses the *public transport and accessibility* collection and the *incidents* collection as its sole data source. The behaviour of these applications is based on this common source: when they capture data, they feed these collections; when they process information, it is obtained from this repository. A correct operation is that of validating the quality of these data.

In summary, the keystone of this architecture is the datastores in the Semantic Subsystem: these datasets (the aforementioned public transport and accessibility collection, and the incidents collection) and their contents are simultaneously the *main result* (the *core*) of the architecture and its extensions (the apps) and the main *source of their behaviour*. The ultimate goal of this architecture is to capture and store this information regarding transport, accessibility and incidents. Once these data have been correctly captured and processed, the apps in the architecture are useful and behave as expected. One of the apps also provides the information that is required by one of the others. Even when all the apps are removed, the architecture is still useful and fulfils a purpose, because it contains the relevant data in the semantic datastore.

The validation of the architecture is, therefore, directly related to the quality of those data. In order to validate the architecture, we intend to verify whether it provides the infrastructure that is able to guarantee (or at least to maintain) the *quality attributes* that sustain these data.

Data quality is, therefore, often described in terms of the conjunction and composition of these attributes, namely *accuracy, completeness, accessibility, redundancy, readability, consistency, usefulness and trust* [9]. The best way in which to validate these attributes is by means of a set of case studies, as a single case study might not be sufficient to describe all of them; and each of the apps consequently provides a case study. This also makes it possible to specifically check certain attributes, considered in isolation from the others. The melange of these relative validations serves to validate the full architecture.

A. DIMENSIONS IN THE ARCHITECTURE

The architecture itself is an abstract construction, signifying that it cannot be directly validated. But as described previously, this integration architecture has been implemented in the domain as the Services *on the move* for Urban mobility (SofUR) platform. Our approach in this section is, therefore, to consider the SofUR system as a ''SoMA in practice''. A full assessment of the platform would, of course, require many details, and these depend on specific technologies (described in Section 3) and on the implementation itself. Our validation will, however, focus on the aspects derived from the *architecture* itself and from its data integration features; the implementation details are of a less importance here.

The validation of the SofUR platform will, therefore, be considered in two different dimensions, namely (1) from a global perspective of *data integration*, as defined by the

SoMA architecture; and (2) from the local point of view provided by each *specific application* on the platform. In fact, the interplay between *global* and *local* viewpoints provides the same kind of assessment already mentioned when discussing *data quality* and its specific *attributes*.

With regard to the first dimension, there are no appropriate data integration metrics at the overall system level; some internal metrics could provide certain insights into the structure of data, but not into their influence on the citizens' welfare. In order to validate this kind of outcome, we must consider some external metrics, i.e. those properties that have a significant influence from a global perspective.

Two of these metrics are the degree of integration (from the user's point of view) and their accessibility. In order to improve the second, and to be able to evaluate the first, it is important to note that the integrated dataset of urban mobility information (semantically annotated with the MANTO ontology [20]) is accessible for download from the platform itself [16], and has been for years, while still continuously growing. Raw data is accessible in the form of Linked Open Data and can, therefore, be processed, and includes information that has been made available only in non-semantic formats, such as GTFS [18].

In addition to this raw semantic (LOD) form, the SofUR platform makes it possible to query these data in a variety of ways. First, the platform provides a Fuseki semantic search engine on top of the Apache Jena server, and is essentially able to answer any potential SPARQL queries. This means that any users can tailor this semantic interface to their particular needs. Moreover, in order to facilitate this kind of queries, the platform itself provides a series of suitable web applications, most remarkably the MMAR and MMSemQuery applications, both of which are focused on the Madrid subway network.

Apart from these semantic interfaces, the platform provides specific means to query specific information in its suite of mobile and web applications, particularly in certain domains. The usefulness of these queries is sometimes defined as the *integration* of data originating from different sources; for instance, in the restricted mobility context, accessibility data are often combined with Open Street Map representations, which serve to locate them and are tightly correlated with them (e.g. in the *MMA4A* app).

Furthermore, some specific queries serve to *separate* information, and consider some data in isolation; for instance, incidents and events also need to be integrated and semantically correlated with their context in the transport network, but these are *distinct* data, with specific features, such as their temporal relevance, and hence have their own processing module and their own repository. However, they are still integrated into the architecture, and are essential in order to keep data updated, by synchronizing their occurrences with the *stable* structure, while defining the current *status*.

The SofUR platform provides different kinds of data processing, with different properties, and its main feature is not the performance level that it attains in them, but the fact that they are *semantically* combined and temporally synchronised; and once they have been *integrated*, any particular application can be provided with them, with the same effort that a local store would require, but with a much richer context.

Even when disregarding their particular origins, once integrated, data have three notorious features in the SofUR platform, namely:

- They are *semantic data*, in the triple sense that they have been semantically annotated, have semantic relationships and define a semantic context for any new data;
- They are *smart data*, as they are not only data gathered from specific selected sources, but they have been processed, harmonised and normalised. This processing ensures (1) basic data *quality*, as any unreliable information is filtered out; (2) *analytic* services, which depend more on relationships than on the volume of data; and (3) the capability to *actively* respond to specific questions and problems using a variety of interfaces.
- They are, finally, *open data* again, as the results (and even crude data) are made openly available in order to enable their downloading and reuse.

The composition of these three features, notably the semantic capabilities, makes a unique combination, and this is in itself one of the values to consider when assessing this platform.

With regard to the second dimension, the impact that this integration architecture has on the citizens' welfare can be more easily evaluated by considering the *utility* of every single application on the SofUR platform. Indeed, apart from making it grow organically, driven by people's needs, the specific applications define the convenience of the platform itself – in addition to the abstract benefits of the data integration (as described above), each specific app provides a particular *utility*. These applications, whether mobile or web, determine the *value* of the platform implemented and the architecture it encloses. In fact, the applications in this suite can also be described as *a set of case studies*, which together assess the validity of the platform.

B. APPLICATIONS AS CASE STUDIES

A brief discussion of the kind of validation provided by these applications, whose enumeration is ordered in a strictly temporal manner, is provided below. Each discussion is related to the data quality attributes listed at the beginning of this section at the end of this summary.

- *Notify.me*, the bus route definition app [38], was the first application in the suite, and it already defined and required the initial version of our architecture. In order to obtain the bus network information, it had to access a variety of sources; in order to compute the routes, it had to use central resources; and it already provided the integrated information in the form of a map. It, of course, has an obvious utility; in practice, calculating transport routes is the first manifestation of *urban computing*.

Moreover, this app was used by the EMT bus company itself, before developing its own version.

- The first app with actual data integration, *MMAppcessible*, also had a clear utility. It already gathered accessibility information – mostly through scraping techniques [20], and it then integrated it with pre-existing data stored in the repository and computed mobility-aware transport routes, with a special emphasis on impaired and blind people, and even had a voice interface. The social significance of this application is, therefore, obvious. Moreover, this version already required a full implementation of the SoMA architecture, which must be considered as the first version of the SofUR platform.
- The next app, *MMA4A*, is an evolution of the above, with even more emphasis being placed on accessibility 16]. It has a better usability, an evolved interface, and a completely expanded architecture. The idea is to add temporary information (incidences, events) to the route computation algorithm. This has three consequences: first, the application itself has to be able to capture these incidences, and must define the interface in order to allow the user to introduce them; second, the architecture must be able to process this crowdsourcing information, accommodating the resulting data stream without overflowing – which is the origin of the smart data subsystem; and third, the semantic subsystem has to be able to synchronise these data and to integrate them with the rest of the repository, achieving a remarkable level of integration, as mentioned previously. The *utility* of this version is significantly expanded with respect to the above version, as it achieves a capability that is rarely found in the context of mobility, i.e. the possibility of avoiding local incidents that compromise the accessibility of a potential valid route, which has to be temporarily circumvented.
- Once the central architecture had a stable structure, the remaining applications were simply built onto it and, therefore, benefit from its contents. The next app, *Access 'n' Go!*, is described in [17], which explores the context of crowdsourcing, discussing what motivates users to provide the information required – on this occasion in the form of a serious game. Our initial experiments with this approach have achieved a significant amount of success and are carefully summarised in [17]. In the first iteration, a group of 20 people got involved in the game in order to capture (and confirm) the accessibility elements in the Madrid subway network, and were able to capture more than 500 of these elements in more than 1450 interactions.

  The web application, *MMAR*, which is still under development, takes advantage of the unique capabilities of a modern web interface to expand beyond the limits of a mobile app. Its purpose is quite similar to that of *MMA4A*, but rather than focusing on providing routes on the move, it is able to exploit the data contained in the repository in a more flexible manner – as described in the first part of this section.
- The semantic web application *MMSemQuery* similarly supplies a web-based interface in order to consume the semantic information contained in the triplestore. As noted above, the main purpose of this application is again to provide an adaptable interface so as to avoid the need to directly access the semantic repository via Fuseki.
- Finally, the mobile app *AccessPedRo*, which is also under development, takes a different approach, by exploiting the same information in order to define *pedestrian routes* on a street map of the city [78]. These routes have been conceived in order to complement the transport network, but have a set of distinct features which, despite the very similar interface, require quite different processing – and the SoMA architecture has, in fact, been expanded to accommodate a specific datastore in order to optimise its reusability. The utility of this application is, when considered in isolation, an additional functionality, directed towards a different group of interest; but when combined with the existing mechanisms, it supplies the means to define a full connectivity mesh on the streets and in the transport network whose significance is yet to be explored.
- Moreover, there are a number of additional models, still under initial development, which will expand the architecture by extending the available functionality with supplementary features. For instance, one module is already able to process the semantic information as it arrives in the repository, and to provide the required normalisation in a principled manner, while another is working to improve the method employed to capture open non-processable information by means of enhanced scraping techniques.

As mentioned previously, these applications also serve as specific case studies, which allow us to specifically study the data quality attributes mentioned. Each of these case studies targets at least one, and possibly more, of these attributes; the intention is to cover them all by combining the specific applications. The SoMA architecture is, of course, still responsible for the *integration dimension* mentioned above, and provides global consistency.

Table 2 shows the correspondence between every single app (case study) and the data quality attributes. Note that the full set is covered by composition, and that the integration architecture plays a central role. As the results of every validation have been positive, we have to conclude that they provide a positive validation of the complete SofUR platform, and the SoMA abstract architecture, which supports their conception and definition.

It should be emphasized that this is not just an assortment of web and mobile projects in a certain domain. We have described the history of the project in order to show how every app plays a logical role in the evolution of the platform. Beyond that, this evolution has also been guided by the need

TABLE 2. Qualitative Impact of Data Attributes on the Different Applications (Case Studies) and Global Impact on the SoMA Architecture.

	Notify.me	*MMAppcessible*	*MMA4A*	*Access'n'Go!*	*MMAR*	*MMSemQuery*	*AccessPedrRo*	*SoMA*
Accuracy	✓	✓	✓	✓	✓	✓	✓	✓
Completeness		✓	✓	✓	✓	✓		✓
Accessibility	✓	✓	✓	✓	✓		✓	✓
Redundancy				✓				✓
Readability	✓	✓	✓	✓	✓	✓	✓	✓
Consistency			✓		✓	✓	✓	✓
Usefulness	✓	✓	✓	✓	✓		✓	✓
Trust	✓			✓		✓		✓

to validate the architecture. For instance, in the initial stages of the platform, it was difficult to *update* information in the repository. Updating obviously affects *consistency*, which is still one of the most significant data quality attributes (including the recent, and relevant, body of work on eventual consistency). In order to validate this consistency we had to be able to update the data within the architecture in a flexible manner. The *MMA4A* case study came into being as a means to provide and test the update capabilities in the architecture, and it is now one of our most successful efforts.

In summary, all of these applications show the wide range of possibilities provided by this semantically integrated architecture, the way in which it is able to organically grow and increase its objectives, and the advantages that it already provides in order to improve the users' well-being in the urban environment. The utility of this approach has, therefore, been shown, and its significance has been validated by describing its existing implementation.

VII. CONCLUSION AND AREAS FOR IMPROVEMENT

The concept of the smart city includes the use of software solutions to meet the challenges involved in improving citizens' daily lives. Urban mobility is a key factor and any improvement to public transport, therefore, provides a benefit for the urban community. On the one hand, traffic jams, energy consumption and carbon emissions will be reduced, and on the other, new parking infrastructures will not be necessary, thus liberating new spaces for pedestrians. All of these changes will undoubtedly increase the welfare of society.

In this context, data captured from hardware sensors, open and public sources, and smart devices must be synchronised and processed in order to perceive the current situation of the city, even in real time. Only then will we be able to address the corresponding solutions in order to improve this situation. This must be done by taking many available sources into account, including their distinct features. The first sources are those provided as *open data*, many using a Linked Open Data structure, ranging from automatic data streams to more elaborate and even pre-processed data sources. There are many other sources that are not strictly *open* or even processable, which supply information in the form of an API (application programme interface). It is, however, probable that the largest potential source of information is not provided by available shared data, but rather by the presence of computational devices in the city, specifically the citizens' *smartphones* and other smart devices. When the citizens themselves are involved, a *crowdsourcing* initiative is defined, and each new software application embodies a different need. As noted previously, a crowdsourcing approach of this nature is required for the success of any smart city enterprise, and should be considered as a vital part of the *urban computing* approach. Processing should also include data already stored in large databases, which define the basic backbone of the city's structural information. In summary, our work proposes the *integration* of even a subset of these data, and we assert that any combination has the potential to benefit the well-being of the individuals affected to a tremendous extent. We have, therefore, defined the following features for this purpose:

- A data integration architecture and platform on which data from these different sources are stored, *semantically* annotated and harmonised, thus ensuring a feature rich combination. As described previously, the platform is organised as three different subsystems (Open Data & Acquisition, Semantic and Smart Data); and
- A suite of software applications designed for both smartphones and the web, and to solve specific issues (accessibility, incidents and infrastructure) one at a time, proving that even small benefits have a significant impact on the global ecosystem.

Our goal has been to provide a generic architecture with which to support the integration of data into the urban mobility platform, serving as the basis for several different applications. Each of them is designed to *improve the well-being* of urban citizens, and to *enrich* the platform itself, increasing its own aggregate intelligence in order to make more sophisticated decisions. As noted above, the key aspect as regards providing this generic integration is the use of semantic technologies.

With regard to future work in this area, we intend to define a systematic and generalised web scraping method that will be based on an underlying reference data model, currently under development, which is directly related to the data found on websites. We also plan to apply generic data processing technologies (such as RML mappings) to these raw collected data, thus providing a unified means to normalise data and transform them into their native semantic (RDF) format.

In summary, our work continues in both directions, and any improvement made to any particular detail will now have a global impact on the whole architecture.

REFERENCES

[1] *AccessMap*. Accessed: Nov. 2020. [Online]. Available: https://www.accessmap.io

[2] *Access2All, EU FP7 Project, Mobility Schemes Ensuring Accessibility of Public Transport for All Users*. Accessed: Nov. 2020. [Online]. Available: https://cordis.europa.eu/project/rcn/89881/factsheet/en

[3] V. Albino, U. Berardi, and R. M. Dangelico, ''Smart cities: Definitions, dimensions, performance, and initiatives,'' *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015, doi: 10.1080/10630732.2014.942092.

[4] *Apache Jena (N.D.)* Accessed: Nov. 2020. [Online]. Available: https://jena.apache.org/index.html

[5] *Apache Kafka (N.D.)* Accessed: Nov. 2020. [Online]. Available: https://kafka.apache.org/intro

[6] *Apache NiFi*. Accessed: Nov. 2020. [Online]. Available: https://nifi.apache.org/

[7] *Apache Tomcat*. [Online]. Available: http://tomcat.apache.org/

[8] Alliance for Telecommunications Industry Solutions. (Mar. 2018). *ATIS Smart Cities. Data Sharing Framework Report*. [Online]. Available: https://www.atis.org/smart-cities-data-sharing/smart-cities-data-sharing.pdf

[9] C. Batini and M. Scannapieco, *Data and Information Quality: Dimensions, Principles and Techniques*. Cham, Switzerland: Springer, 2016, p. 23.

[10] *Beautiful Soup*. Accessed: Nov. 2020. [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

[11] T. Berners-Lee, R. Fielding, and L. Masinter. (2005). *Uniform Resource Identifier (URI): Generic Syntax*. [Online]. Available: http://www.hjp.at/doc/rfc/rfc3986.html#sec_1.1

[12] S. Bischof, A. Polleres, and S. Sperl, ''City data pipeline,'' in *Proc. I-SEMANTICS Posters Demonstrations Track*, 2013, p. 45.

[13] D. Brickley and R. V. Guha. (2014). RDF schema 1.1. W3C Recommendation. [Online]. Available: https://www.w3.org/TR/rdf-schema

[14] *BUSUP*. Accessed: Nov. 2020. [Online]. Available: https://www.busup.com/

[15] *Ciudades Patrimonio de la Humanidad*. Accessed: Nov. 2020. [Online]. Available: http://www.ciudadespatrimonio.org/accesibilidad/?idioma=en

[16] P. Cáceres, C. E. Cuesta, A. Sierra-Alonso, B. Vela, J. M. Cavero, and M. A. Garrido, ''Even smarter data: Using crowdsourcing to improve accessibility in real-time,'' in *Proc. SMART ACCESSIBILITY*, Athens, Greece, 2019, pp. 28–32.

[17] P. Cáceres, C. E. Cuesta, B. Vela, J. M. Cavero, and A. Sierra, ''Smart data at play: Improving accessibility in the urban transport system,'' *Behav. Inf. Technol.*, vol. 39, no. 6, pp. 681–694, Jun. 2020, doi: 10.1080/0144929X.2019.1652852.

[18] 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 *Proc. 2nd Int. Conf. Universal Accessibility Internet Things Smart Environ. (SMART ACCESSIBILITY)*, Nice, France, 2017, pp. 6–9.

[19] P. Cáceres, A. Sierra-Alonso, C. E. Cuesta, B. Vela, and J. M. Cavero, ''Modelling and linking accessibility data in the public bus network,'' *J. Universal Comput. Sci.*, vol. 21, no. 6, pp. 777–795, 2015.

[20] P. Cáceres, A. Sierra-Alonso, B. Vela, J. M. Cavero, and C. E. Cuesta, ''Towards smart public transport data: A specific process to generate datasets containing public transport accessibility information,'' in *Proc. 3rd Int. Conf. Universal Accessibility Internet Things Smart Environments (SMART ACCESSIBILITY)*, Rome, Italy, 2018, pp. 66–71.

[21] P. Cáceres, A. Sierra-Alonso, B. Vela, J. M. Cavero, M. A. Garrido, and C. E. Cuesta, ''Normalizing public transport accessibility data in real-time,'' *Int. J. Adv. Intell. Syst.*, vol. 12, nos. 3–4, pp. 191–202, 2020.

[22] P. Cáceres, A. Sierra-Alonso, B. Vela, J. M. Cavero, M. A. Garrido, and C. E. Cuesta, ''Adding semantics to enrich public transport and accessibility data from the Web,'' *Open J. Web Technol.*, vol. 7, no. 1, pp. 1–18, 2020.

[23] C. Cardonha, D. Gallo, P. Avegliano, R. Herrmann, F. Koch, and S. Borger, ''A crowdsourcing platform for the construction of accessibility maps,'' in *Proc. 10th Int. Cross-Disciplinary Conf. Web Accessibility (W4A)*, 2013, pp. 1–4, doi: 10.1145/2461121.2461129.

[24] CIVITAS Initiative. (2012). *Cleaner and Better Transport in Cities*. [Online]. Available: http://www.civitas.eu

[25] *CKAN*. Accessed: Nov. 2020. [Online]. Available: https://ckan.org/

[26] J. Clark and S. DeRose. (2016). XML path language (XPath). W3C Recommendation. [Online]. Available: https://www.w3.org/TR/1999/REC-xpath-19991116/

[27] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl, ''Understanding smart cities: An integrative framework,'' in *Proc. 45th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2012, pp. 2289–2297.

[28] C. E. Cuesta, P. Cáceres, B. Vela, and J. M. Cavero, ''CoMobility: A mobile platform for transport sharing,'' *ERCIM News*, no. 93, pp. 22–23, 2013.

[29] R. Cyganiak, D. Wood, and M. Lanthaler. (2014). RDF 1.1: Concepts and abstract syntax. W3C Recommendation. [Online]. Available: https://www.w3.org/TR/rdf11-concepts/.

[30] F. Darari and R. Manurung, ''LinkedLab: A linked data platform for research communities,'' in *Proc. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, 2011, pp. 253–258.

[31] A. Dimou, M. V. Sande, P. Colpaert, R. Verborgh, E. Mannens, R. Van de Walle, ''RML: A generic language for integrated RDF mappings of heterogeneous data,'' in *Proc. 7th Workshop Linked Data Web*, 2014, pp. 1–5.

[32] *EMT Madrid Home Page*. Accessed: Nov. 2020. [Online]. Available: http://www.emtmadrid.es/Index.aspx?lang=en-GB

[33] B. Ferris, K. Watkins, and A. Borning, ''OneBusAway: Results from providing real-time arrival information for public transit,'' in *Proc. CHI*, Atlanta, Georgia, Apr. 2010, pp. 1807–1816.

[34] *Fuseki*. Accessed: Nov. 2020. [Online]. Available: http://jena.apache.org/documentation/fuseki2/

[35] S. Gaggi, T. Fluhrer, and T. Janitzek. Innovation in urban mobility: Policy making and planning. Directorate-general for mobility and transport. European Union, Luxembourg. [Online]. Available: https://www.kowi.de/Portaldata/2/Resources/fp/tripurban-mobility.pdf

[36] F. Gandon and G. Schreiber. (2014). RDF 1.1 XML syntax. W3C Recommendation. [Online]. Available: https://www.w3.org/TR/rdf-syntax-grammar/

[37] F. Gao, M. I. Ali, and A. Mileo, ''Semantic discovery and integration of urban data streams,'' in *Proc. 5th Workshop Semantics Smarter Cities*, 2014, pp. 15–30.

[38] M. A. Garrido, P. Cáceres, B. Vela, C. E. Cuesta, J. M. Cavero, and A. Sierra-Alonso, ''Consuming Web data in a guiding app for public bus users,'' *Open J. Web Technol.*, vol. 5, no. 1, pp. 31–43, Sep. 2018.

[39] M. A. Garrido and P. Cáceres, ''A data-driven methodology to generate living ontologies,'' *Appl. Ontol.*, to be published.

[40] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanović, and E. Meijers. (2007). Smart cities: Ranking of European medium-sized cities. Centre of Regional Science (SRF). Vienna University of Technology, Vienna, Austria. [Online]. Available: http://www.smart-cities.eu/download/smart_cities_final_report.pdf

[41] GTFS. *General Transit Feed Specification*. Accessed: Nov. 2020. [Online]. Available: https://developers.google.com/transit/gtfs/

[42] G. L. Hajba, *Website Scraping With Python: Using BeautifulSoup and Scrapy*. New York, NY, USA: APress, 2018.

[43] R. E. Hall, ''The vision of a smart city,'' in *Proc. 2nd Int. Life Extension Technol. Workshop*, Paris, France, 2000.

[44] C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, and P. Williams, ''Foundations for smarter cities,'' *IBM J. Res. Develop.*, vol. 54, no. 4, pp. 1–16, Jul. 2010.

[45] S. J. Humer and E. Foster, ''A comparative analysis of the C++, Java, and Python languages,'' Princ. Program. Lang., Keene State College, New Hampshire, OH, USA, Tech. Rep., 2014.

[46] *Identification of Fixed Objects in Public Transport (IFOPT), CEN/TC 278 Intelligent Transport Systems*, Public Transport Standard EN 28701, 2012.

[47] H. Jeekel, ''Social sustainability and smart mobility: Exploring the relationship,'' *Transp. Res. Procedia*, vol. 25, pp. 4296–4310, Jan. 2017.

[48] D. Kouzis-Loukas, *Learning Scrapy*. Birmingham, U.K.: Packt Publishing, 2016.

[49] S. L. Lau and S. M. S. Ismail, ''Towards a real-time public transport data framework using crowd-sourced passenger contributed data,'' in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC-Fall)*, Sep. 2015, pp. 1–6.

[50] X. Liu, A. Heller, and P. S. Nielsen, ''CITIESData: A smart city data management framework,'' *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 699–722, Dec. 2017, doi: 10.1007/s10115-017-1051-3.

[51] P. Lombardi, S. Giordano, H. Farouh, and W. Yousef, "Modelling the smart city performance," *Innov., Eur. J. Social Sci. Res.*, vol. 25, no. 2, pp. 137–149, Jun. 2012.

[52] V. Lopez, S. Kotoulas, M. L. Sbodio, M. Stephenson, A. Gkoulalas-Divanis, and P. M. Aonghusa, "QuerioCity: A linked data platform for urban information management," in *The Semantic Web* (Lecture Notes in Computer Science), vol. 7650. Berlin, Germany: Springer-Verlag, 2012, pp. 148–163.

[53] M. Lytras and A. Visvizi, "Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research," *Sustainability*, vol. 10, no. 6, p. 1998, Jun. 2018, doi: 10.3390/su10061998.

[54] M. Lytras, A. Visvizi, and A. Sarirete, "Clustering smart city services: Perceptions, expectations, responses," *Sustainability*, vol. 11, no. 6, p. 1669, Mar. 2019, doi: 10.3390/su11061669.

[55] (2011). *Mediate, EU FP7 Project, Methodology for Describing the Accessibility of Transport in Europe*. [Online]. Available: https://trimis.ec.europa.eu/

[56] *Metro Madrid Home Page*. Accessed: Nov. 2020. [Online]. Available: http://www.metromadrid.es/en/

[57] R. Mitchell, *Web Scraping With Python: Collecting Data from the Modern Web*. Sebastopol, CA, USA: O'Reilly Media, 2015.

[58] *MongoDB*. Accessed: Nov. 2020. [Online]. Available: https://www.mongodb.com/what-is-mongodb

[59] *Moovit*. Accessed: Nov. 2020. [Online]. Available: https://moovit.com/

[60] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," in *Proc. 12th Annu. Int. Digit. Government Res. Conf. Digit. Government Innov. Challenging Times*, College Park, MD, USA, Jun. 2011, pp. 282–291.

[61] B. Ning, H.-R. Dong, D. Wen, L. Li, and C.-J. Cheng, "ACP-based control and management of urban rail transportation systems," *IEEE Intell. Syst.*, vol. 26, no. 2, pp. 84–88, Mar. 2011.

[62] E. Prud'hommeaux and A. Seaborne. (2008). SPARQL 1.1 query language for RDF. W3C Recommendation. [Online]. Available: https://www.w3.org/TR/rdf-sparql-query/

[63] *OpenTRipPlanner*. Accessed: Nov. 2020. [Online]. Available: http://www.opentripplanner.org/

[64] *OpenStreetMap*. Accessed: Nov. 2020. [Online]. Available: https://www.openstreetmap.org

[65] D. Puiu, P. Barnaghi, R. Tönjes, D. Kümper, M. I. Ali, A. Mileo, J. X. Parreira, M. Fischer, S. Kolozali, N. Farajidavar, F. Gao, T. Iggena, T.-L. Pham, C.-S. Nechifor, D. Puschmann, and J. Fernandes, "CityPulse: Large scale data analytics framework for smart cities," *IEEE Access*, vol. 4, pp. 1086–1108, 2016, doi: 10.1109/ACCESS.2016.2541999.

[66] *Python*. Accessed: Nov. 2020. [Online]. Available: https://docs.python.org/3

[67] *Python Documentation*. Accessed: Nov. 2020. [Online]. Available: https://docs.python.org/3/library/urllib.request.html?highlight=requests%20library

[68] RDF Working Group, W3C Recommendations. (2014). *Resource Description Framework (RDF)*. [Online]. Available: http://www.w3.org/RDF/

[69] *Request Library*. Accessed: Nov. 2020. [Online]. Available: https://2.python-requests.org

[70] H. Santos, P. Pinheiro, and D. L. McGuinness, "Contextual data collection for smart cities," in *Proc. 6th Workshop Semantics Smarter Cities*, 2015, pp. 53–68.

[71] T. Sarjakoski, P. Kettunen, H.-M. Halkosaari, M. Laakso, M. Rönneberg, H. Stigmar, and T. Sarjakoski, "Landmarks and a hiking ontology to support wayfinding in a national park during different seasons," in *Cognitive and Linguistic Aspects of Geographic Space* (Lecture Notes in Geoinformation and Cartography), M. Raubal, D. M. Mark, and A. U. Frank, Eds. Berlin, Germany: Springer, 2013, pp. 99–119, doi: 10.1007/978-3-642-34359-9_6.

[72] J. Schlingensiepen, R. Mehmood, and F. C. Nemtanu, "Framework for an autonomic transport system in smart cities," *Cybern. Inf. Technol.*, vol. 15, no. 5, pp. 50–62, Apr. 2015, doi: 10.1515/cait-2015-0016.

[73] *Scrapy*. Accessed: Nov. 2020. [Online]. Available: https://scrapy.org/

[74] (2019). *Spring Boot*. [Online]. Available: http://spring.io/projects/spring-boot

[75] *Swiftly*. Accessed: Nov. 2020. [Online]. Available: https://www.goswift.ly/

[76] *Road Transport and Traffic Telematics Public Transport*, Transmodel Standardization, E C Transmodel, Ref. Data Model, Centre d'Études Techniques de l'Équipement Méditerrannée (CETE), Aix-en-Provence, France, 2015.

[77] E. J. Tomaszewska and A. Florea, "Urban smart mobility in the scientific literature—Bibliometric analysis," *Eng. Manage. Prod. Services*, vol. 10, no. 2, pp. 41–56, Jun. 2018, doi: 10.2478/emj-2018-0010.

[78] B. Vela, J. M. Cavero, P. Caceres, A. Sierra, and C. E. Cuesta, "Defining a NoSQL document database of accessible transport routes," in *Proc. IEEE Int. Conf. Internet Things (iThings), IEEE Green Comput. Commun. (GreenCom), IEEE Cyber, Phys. Social Comput. (CPSCom), IEEE Smart Data (SmartData)*, England, U.K., Jun. 2017, pp. 1125–1129.

[79] B. Vela, P. Cáceres, J. M. Cavero, A. Sierra, C. E. Cuesta, and M. A. Garrido, "Guíame EMT," Spanish Patent M-004237/2014, Nov. 18, 2014.

[80] B. Vela, J. M. Cavero, P. Caceres, and C. E. Cuesta, "A semi-automatic data–scraping method for the public transport domain," *IEEE Access*, vol. 7, pp. 105627–105637, 2019, doi: 10.1109/ACCESS.2019.2932197.

[81] *VIVO Application*. Accessed: Nov. 2020. [Online]. Available: https://duraspace.org/vivo/

[82] D. Washburn, U. Sindhu, S. Balaouras, R. A. Dines, N. M. Hayes, and L. E. Nelson. (2010). Helping CIOs understand 'smart city' initiatives: Defining the smart city, its drivers, and the role of the CIO. Forrester Research, Cambridge, MA, USA. [Online]. Available: http://public.dhe.ibm.com/partnerworld/pub/smb/smarterplanet/forr_help_cios_und_smart_city_initiatives.pdf

[83] *Wheelmap*. Accessed: Nov. 2020. [Online]. Available: https://wheelmap.org

[84] *Wheelmate*. Accessed: Nov. 2020. [Online]. Available: http://myhealthapps.net/app/details/13/wheelmate

[85] J. Zimmerman, A. Tomasic, C. Garrod, D. Yoo, C. Hiruncharoenvate, R. Aziz, N. R. Thiruvengadam, Y. Huang, and A. Steinfeld, "Field trial of tiramisu: Crowd-sourcing bus arrival times to spur codesign," in *Proc. CHI*, New York, NY, USA, 2011, pp. 1677–1686.

PALOMA CÁCERES received the M.Sc. degree in computer science from the Polytechnic University of Madrid, Spain, in 1993, and the Ph.D. degree in computer science from Universidad Rey Juan Carlos, Madrid, Spain, in 2006. She is currently an Associate Professor with Universidad Rey Juan Carlos. Her research interests include software engineering, web engineering, model driven development, data science, open and linked data, semantic Web, and transport and accessibility. She has authored or coauthored several national and international articles related to these areas.

ALMUDENA SIERRA-ALONSO received the Ph.D. degree in computer science from the Universidad Politécnica de Madrid, in 2000. She is currently an Associate Professor with the Computer Science and Statistics Department, Universidad Rey Juan Carlos. She has authored or coauthored several articles related to software engineering, semantic web, and teaching in engineering education (in areas such as operating systems and software engineering). Her research interests include software engineering, data science, open and linked data, semantic Web, and transport and accessibility.

CARLOS E. CUESTA received the Ph.D. degree in information technologies from the University of Valladolid, in 2002. He is currently an Associate Professor of software engineering with Universidad Rey Juan Carlos, Madrid, Spain. He has published in premier conferences and journals, such as the *International Journal of Information Technology & Decision Making*, and *Future Generation Computer Systems*. His current research interests include software architecture, including such topics as self-adaptive systems and systems-of-systems, and with relation to concurrency theory, formal methods, distributed systems, and data engineering. He has been the Program Chair of the 6th and 12th European Conferences on Software Architecture (ECSA 2012 & 2018).

BELEN VELA received the M.Sc. degree in computer science from Carlos III University and the Ph.D. degree in computer science from Universidad Rey Juan Carlos, Madrid, Spain. She belongs to the Vortic3 Research Group. She is currently an Associate Professor with the Department of Computing Science, Computer Architecture, Programming Languages and Systems and Statistics and Operative Investigation, Universidad Rey Juan Carlos. She has participated in and led several research projects and has had numerous articles published in prestigious journals and conferences. Her research interests include software engineering, information system engineering, data science, open data, databases (NoSQL), model driven development, transport, accessibility, and scientometrics.

JOSÉ MARÍA CAVERO BARCA received the M.Sc. degree in computer science from the Polytechnic University of Madrid and the Ph.D. degree in computer science from Universidad Rey Juan Carlos. He has been the Dean of the School of Computer Science Engineering for four years. He is currently an Associate Professor with the School of Computer Science Engineering, Universidad Rey Juan Carlos. His research interests include databases, open data, transport, accessibility, and scientometrics.