A Collaborative System to Manage Information Sources Improving Transport Infrastructure Data Knowledge

Salvatore Cafiso, Alessandro di Graziano & Giuseppina Pappalardo*

Department of Civil Engineering and Architecture, University of Catania
Viale Andrea Doria 6, I-95125 Catania, Italy
*E-mail: giusy.pap@dica.unict.it

Abstract. The present paper describes the WIKI RoadSMap project implemented within a start-up research program. The main objective of the project is to create a system that applies innovative technologies to information gathered to enable the acquisition of greater local knowledge and analysis of issues related to road infrastructure and directly and indirectly connected elements. By applying semantic analysis technology for the extraction, collection, integration and publication of data, WIKI RoadSMap allows users to acquire greater knowledge in order to optimize choices related to road infrastructure. The system allows more detailed and targeted dissemination of data related to the design, management and maintenance of an infrastructure. The source and type of data needed are different and heterogeneous, including information 'posted' by people with private and/or commercial purposes, or available at road agencies and/or public administrations or related to specific surveys carried out. The system platform should be available on the Web and on smartphones, both providing different levels of access and subscriptions. The spread and use of WIKI RoadSMap could have a positive impact on the market with regard to the supply of materials and specialized technical skills and companies operating in the areas of interest.

Keywords: *infrastructure; open data; road agencies; road users; survey.*

1 Introduction

Governments have large amounts of basic data that can be of economic and social value to society as a whole. Along those lines, more and more European countries are developing policies to release these data as open (government) data. This term refers to information that can be freely used, modified, and shared by anyone for any purpose. It must be available under an open license and provided in a convenient and modifiable form that is machine-readable. Information on road infrastructure is correlated only with the transport topic in terms of traffic data. In this context, the main target of the WIKI RoadSMap project was to design a system that applies innovative technologies for semantic analysis to the information gathered to enable the acquisition of greater local knowledge and analysis of road infrastructure issues and directly and indirectly

Received January 9th, 2019, Revised October 30th, 2019, Accepted for publication November 29th, 2019.
 ISSN: 2337-5779, DOI: 10.5614/j.eng.technol.sci.2019.51.6.8

connected elements. Specifically, the system allows a more detailed and targeted dissemination of processed data relating to the design characteristics of the road infrastructure, such as pavement, geometry, and several management parameters such as traffic flow and road crashes.

Data sources connected to road infrastructure can be: information 'posted' by people for different reasons, also in connection with commercial purposes, such as materials used for specific processes and their performance, traffic limitations, references and standards, what is absolutely not recommended for that type of activity, etc.; information available at road agencies and public administrations directly or indirectly connected to open data policies; technical surveys carried out in order to collect more detailed information, related to significant data for instance, connected to road assets or traffic flow. The heterogeneous nature of the sources and of the data results in an application of models and integration tools that takes advantage of innovative technologies in order to provide meaningful and geo-referenced information and to allow consultation and browsing through maps.

The types of analyzed data also refer to different technical levels, so that it is possible to hypothesize varying levels of access to the database, providing both free access to certain information that is available via the Web and a mechanism of subscriptions for those who want to access more specific technical information. The system should also be accessible via the Web or dedicated applications on mobile devices. In the present study, after a brief description of the state of the art on road infrastructure data availability, the WIKI RoadSMap system is presented in terms of data acquisition and integration, database storage and management, and data analysis and applications.

2 State of the Art

In the literature there are several studies that have analyzed the accessibility of information connected with the infrastructure of an area in order to increase the knowledge directly and indirectly connected to their use, such as materials and construction techniques for the realization and maintenance of infrastructure or traffic information for faster or safer routes. In this context, there are many public and private data sources from which it is possible to obtain information [1]. Among these there are many examples of accessible information in the form of open data, which are particular types of data where access is free and there are no copyright restrictions, patents or other forms of control and limitation of reproduction. In Italy, for example, as early as 2011, the open data produced by public administrations were collected and made available through the portal dati.gov.it. A useful non-government source of open data on the international scene is the OpenStreetMaps project, which gives free access and

usage of images of maps of road networks from all over the world and the possibility to modify them. Among the paths to find information, we can find the WiKi system, which today represents a collaborative way to create information and content via the open contribution by all users who can add, edit or delete content in a simple way and with appropriate instruments.

At the national level it is important to refer to the ARTIST (Telematics Architecture for the Italian Transport System) program [2]. ARTIST is an evolving framework for the creation and deployment of a telematics system for traffic and transportation in multi/intermodal ways. It defines, therefore, the set of necessary services, functional relations, and the key characteristics of logical and physical organizational relationships in order to match existing services with user needs. Contributing to the definition of the models and algorithms for the processing of raw data in order to derive functional information that cannot be detected there, is a vast evolving international body of literature, on which numerous applications are based, including commercial, analysis and simulation software. At the European level, the Etis-agent project [3] has developed an integrated system made up of three components: data integration, data storage and front-end environment.

Data integration refers to those processes through which transactional or operational data (e.g. road tolls) residing in dispersed, heterogeneous data sources is integrated to produce policy-related indicators. Data storage refers to a repository used to store the variables and the policy-related indicators produced from data integration. Data storage is materialized as a data warehouse. The indicators are stored in the data warehouse as multidimensional structures (data cubes), which allow policy makers to view an indicator from a number of policy related perspectives. The front-end environment provides a distinct interface to access, compute and visualize policy-related indicators and their metadata. Policy makers are able to relate policies to indicator metadata browsing and editing, access indicators, and apply online analytical processing operations to them.

3 Data Acquisition and Integration

The existing literature about the use of open database management tools can be categorized according to four types of information that is typically captured though applications and is then disseminated in various ways to be utilized in different applications. The types of data include: traffic, infrastructure, environmental, and behavior information [4]. But they work for each type of information, without the possibility to merge all. For this reason, we propose WIKI RoadSMap as a system to be implemented with the aim of managing a large set and formats of data relating to road infrastructure. It is a collaborative

system that uses an organic and innovative data stream to merge information connected to road infrastructures and their surrounding area, by applying semantic analysis technology for the extraction, collection, integration and publication of data.

Data heterogeneity resulted in the implementation of a tool that takes advantage of innovative technologies in order to integrate information available from public administrations in the form of open data, information 'posted' on the web by users and data obtained from specific surveys. The WIKI RoadSMap structure is shown in Figure 2, where it is highlighted how the database storage represents the core of the entire system.

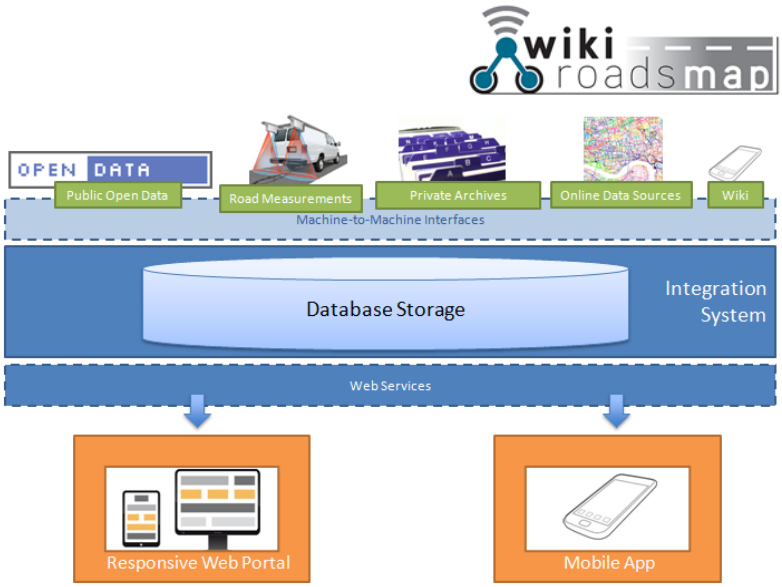

Figure 1 WIKI Roads Map structure.

3.1 Open Data Related to Road Network. Survey On the Italian State of the Art

Data held by public administrations are a huge asset and are becoming increasingly important. In fact, thanks to the use of info-telematic technologies, it is possible to use these data to make the administration more transparent, to provide efficient services, and to apply them in different areas from those for which they were specifically collected. To allow adequate dissemination of the open data, a national portal of open data was developed, in which references of the data opened by each administration are managed. In this context, in order to distinguish the different formats that can be used in the data set coding, a

cataloguing model has been proposed in the W3C (World Wide Web Consortium) [5] that classifies data according to their characteristics on a scale of values from 1 (one star) to 5 (five stars), where 1 refers to the basic model (unstructured, legible, printable files, stored locally on a PC but which cannot be processed) and 5 refers to linked open data (open data that have links to other datasets).

A specific task in the WIKI RoadSMap project highlights the lack of availability of transport infrastructure information in the form of open data that can be used for asset management as identified in the project. For instance, in Italy currently only 402 open datasets connected to the road network (only 3.3% of the 12,000 total) are available, a number that highlights how little interest the administrations have to collect and disseminate data related to the road context. Furthermore, 61% of the datasets are issued by municipal administrations, followed by regional and provincial administrations with 18%, while only the central administration is present on the national Open Data Database. Finally, with reference to the cataloguing proposed in W3C, 50% of the datasets concerning the road context are classified with three stars and only 11% have a high reusability level (4 stars) (Figure 2).

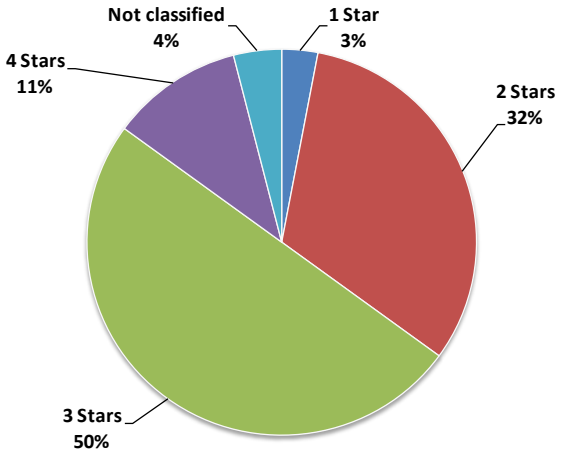

Figure 2 Summary statistics for dataset coding.

3.2 Online Data Source

Specific subsections of open data can originate from information ‘posted’ by people for different reasons, also connected to commercial purposes, all connected with road infrastructure asset management, such as materials used for specific processes and their performance, traffic limitations, references and standards. Today, WiKi systems represent a collaborative way to create

information via open contribution by all users who can add, edit or delete contents in a simple way and with appropriate instruments. WIKI RoadSMap was designed to find information among these paths.

3.3 Data Acquired Through Surveys with Advanced Techniques

Technical information with higher quality can be collected with specific surveys and equipment. These data are often not immediately available, but they could become accessible within a logic of shared valued information. With the aim of analyzing potentialities and problems of integrating the available data with field measurements carried out with advanced equipment, during the project some road infrastructures were surveyed using ARAN 9000. Automatic Road ANalyzed 9000 was recently acquired by the University of Catania, which is carrying out an extensive research program for system calibration and output adaptation to a performance-based inventory of road infrastructures [6-8].

ARAN 9000 is one of the most advanced high-performance systems for road asset monitoring with a productivity of 200 km/day that allows collection of more than fifteen different data streams continuously and at varying capture rates in a single pass at traffic speed. Several system components are integrated into an advanced data collection vehicle equipped with a precision survey system and software able to perform the collection of a road's geometric data (cross section, gradients, horizontal and vertical alignment), pavement unevenness, detection and classification of distress and road asset inventory data including road furniture (e.g. sign, barriers).

All the data recorded are accurately positioned using a geographical reference system with the help of a satellite differential correction receiver (RTK DGPS), and coupled with an inertial platform (IMU) and using a high-resolution odometer (DMI) to provide a sub-meter positioning in every environmental condition because the inertial aided navigation system ensures that GPS accuracy is maintained even during long satellite blackouts (e.g. tunnels) with sub-meter precision. Asset inventory detection is carried out using the HD Videolog system and the Surveyor® asset extraction software.

The high quality of the two front cameras and one rear camera with resolutions of 2750 x 2200 pixels ensure complete detection of the road environment in a 240° field of view (Figure 3). The Videolog and GPS systems work in sync to provide the geo-referencing of digital images. The analysis of images for extracting road inventory assets is carried out by using a software tool that allows to display, measure and inventory any road features displayed in the image (e.g. lane, shoulder width, sign, marking, roadside, safety barriers, etc.).

Figure 3 Road Video Survey.

It is possible to merge ARAN data with data derived from on-site continuous monitoring systems, such as high-speed devices/methods [9,10] and sensors/smartphones [11,12].

4 Database Storage and Management

The technological core of the WIKI RoadSMap infrastructure is an integrated storage system through which it is possible to merge all information coming from the various data sources previously identified. In this context, the decision of the technology to be used and the general architecture has a strategic importance for the success of the project. For this reason, the first phases of the research activity focused on the study of the state of the art systems available and on their comparison, in order to identify a set of tools, technologies and infrastructures that could constitute the general architecture of the system of different WIKI RoadSMap storages.

Given the nature of the information that the database must manage, scouting of solutions and technologies was carried out, with a particular focus on those available under an open source license, oriented to the specific management of data whose size and complexity are such that they require specific tools for archiving, management and visualization. The choice of the different technologies to be evaluated has been oriented in particular to the so-called NoSQL systems (Not Only SQL), since the classic database engines represented by RDBMS (or relational database) systems are not suitable to manage the types and quantities of data characterized by the specific properties of big data.

From the analysis of the technical characteristics of the database, it was possible to identify the most suitable solution for its management in MongoDB [10]. Its high reliability features (in terms of consistency and partition tolerance), enables a powerful network of allowable queries (including Geospatial Queries) and other features such as the wide availability of documentation and literature, the activity and size of the open source community, and the simplicity of implementation and dialogue with most programming languages thanks to the wide availability of interfaces. The infrastructure, designed to host the WIKI RoadSMap storage platform, was built on OpenStack technology and was consisted of a series of virtual machines with a 64-bit Ubuntu Server 12.4 operating system. The architecture of the infrastructure hosting the storage system was designed while taking into account the requirements of (a) security, (b) service availability, and (c) scalability.

From a security point of view, the operating system of virtual machines has been extended with a series of packages that manage firewalling (ufw), access security (VPN), and configuring the permissions of the various users and system administrators. From the point of view of the availability of the service, a server clustering mechanism was implemented that enabled the display of services from two different virtual machines potentially residing on delocalized data centers. Furthermore, an automatic backup mechanism was implemented that sends daily data to an additional external server to allow recovery within a 24-hour recovery point objective.

As for scalability, the choice of using redundant virtual machines increases the storage capacity in a dynamic and proportional manner to the application and storage needs that will come, without affecting the availability of the service. The system was explicitly designed and implemented to be able to collect data from heterogeneous information sources by type (relational/non-relational), information content (meteorological data, traffic data, related to infrastructural features, etc.), and dynamism, in agreement with the standard for the representation of information used by the main geolocation systems (ShapeFile, OSM, GeoJSON, etc.) and related integration and conversion technologies (Figure 4).

In order to respect the uniformity and consistency of information, it was decided to develop a conversion mechanism that allows to harmonize the spatial representation by converting each type of representation into another. To this end, a system was developed that allows the association of GPS points to specific OSM objects that, combined with the conversion system of the data geo-referencing format, allows any geo-referenced information – in any format – to be traced back to a specific stretch of infrastructure in OpenStreet Maps.

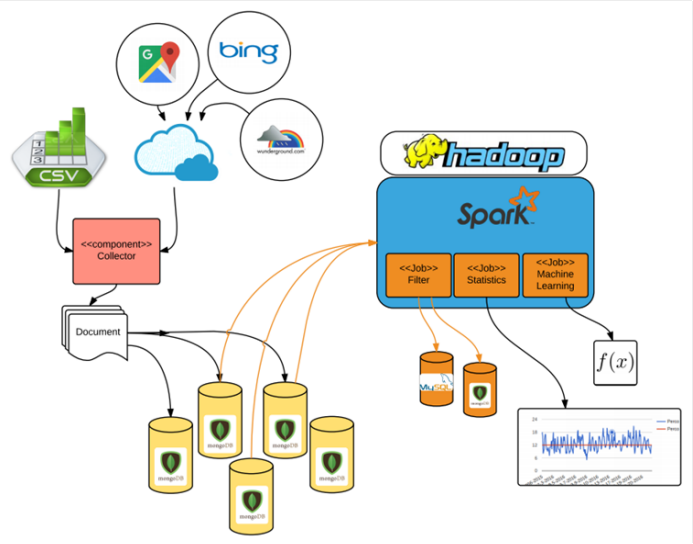

Figure 4 Database storage architecture.

5 Data Analysis and Service

Finally, access and processing system data stored in the Big Data WikiRoads Server database infrastructure were implemented in the form of Web portals and mobile applications: 'WIKI RoadSMap' and 'Wiki Reporter'. The WIKI RoadSMap app, which can be accessed using any Web-enabled browser and mobile device connected to the Internet, allows the information available in the storage system to be viewed on a map through a search that is available in two ways (Figure 5): 1) a search on single points, identified by positioning a placeholder or by identifying the GPS position of the device in use; 2) a search on segments, definable by setting the start and end points, in which case the system automatically calculates the sections of road infrastructure intercepted by the two points, highlighting the segment of interest for research. To get the result, you need to click on the right button that launches the request to the server. The requested data can be filtered by data type through a modal window with a checkbox list from which to select the categories of elements that the user wishes to display on the map (Figure 6).

Markers and road segments are placed on the map with a particular color that groups elements of the same category. Finally, selecting the marker or the segment of interest will open an information window that contains all the informative details of the element in Json format. The Wiki Reporter app allows users to locate a point on the map – eventually the point of their GPS coordinates will be selectable by means of a special button – and allows to

provide information relating to that point by sending a ‘report’. The app works on both iOS and Android devices, adapts to multiple screen sizes, and does not require registration or login.

The user can view the map of the territory with the possibility of zooming and moving, centering the map on the current GPS position taken from the device, and selecting a point on the map to give geo-referenced information on the coordinates of the chosen point. The types of information that can be entered are selectable from a predefined list: Traffic, Weather, Infotraffic, Incidents, Pavement, RMS, RUT, Roughness, Access, and Accident Indicator. The application contacts the storage server of WIKI RoadSMap to send information via the Internet over the http protocol and the user receives feedback confirming the transmission was made.

On a graphical level, the interface design takes advantage of the simplicity and vivacity of a ‘flat design’, which manages to appear less ‘flat’ and more ‘realistic’ by virtue of tactile surfaces. Realism is also accentuated by animations, which in turn support the movements and actions performed by the user. At the level of graphical components, the interface design is characterized by: a GPS activation button, with the ability to quickly center the map displayed to the coordinates of the device in use; an Add Marker button to identify the point to be signaled; Zoom In and Zoom Out buttons; buttons for choosing the type of map to be displayed (map/satellite).

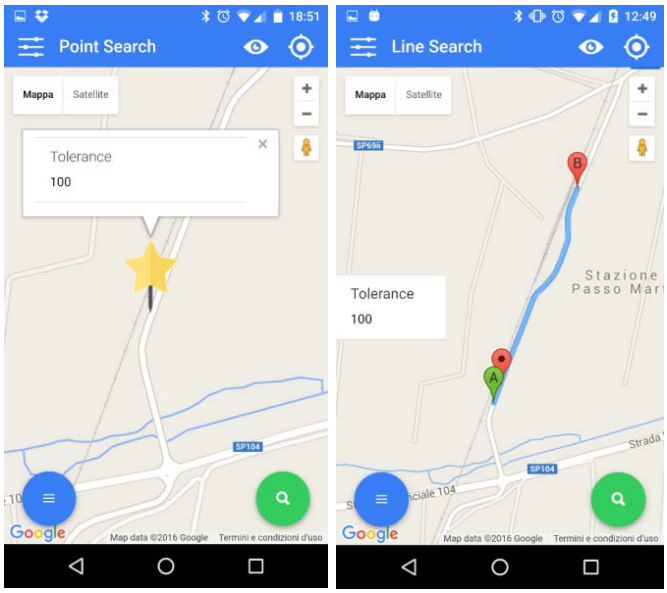

Figure 5 Information view through WIKI RoadSMap app.

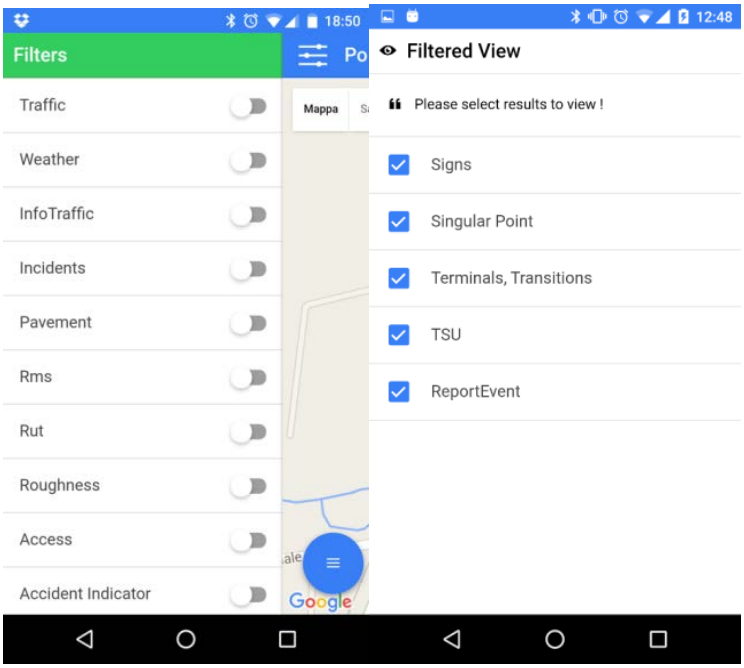

Figure 6 Data request on the WIKI RoadSMap app.

Below (Figure 7) are some preliminary mock-ups for the interface design:

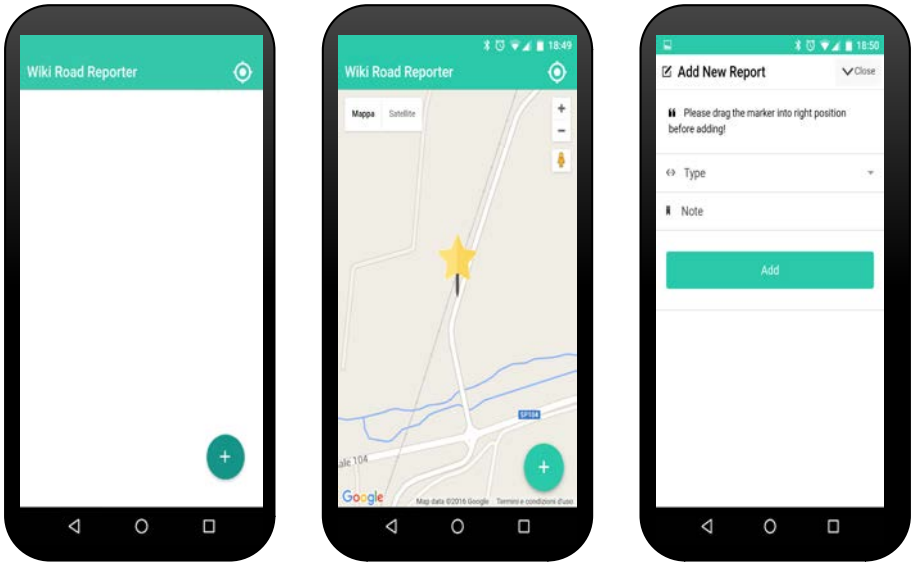

Figure 7 Wiki reporter interface.

6 Conclusions

In this paper, a collaborative system, named WIKI RoadSMap, developed in the framework of a start-up program, is presented. It applies an organic and innovative data-stream that merges information connected with road infrastructure asset monitoring and management. WIKI RoadSMap uses a semantic analysis technology for the extraction, collection, integration and publication of data, allowing users, both public and private, to increase knowledge directly and indirectly connected to road infrastructure for different purposes.

The technological heart of the WIKI RoadSMap is an integrated storage system through which it is possible to merge all the information coming from the various data sources, such as open data, Wiki Web-free openly editable platforms and technical surveys. The architecture of the system was built on OpenStack technology and was designed taking into account the requirements of: (a) security, (b) service availability, and (c) scalability. The system was explicitly designed and implemented in order to merge data coming from heterogeneous information sources by type (relational/non-relational), content (meteorological data, traffic data, features, etc), geo-location systems (ShapeFile, OSM, GeoJSON, etc.) and related conversion technologies. User processing systems were implemented in the form of Web portals and mobile applications, called WIKI RoadSMap and Wiki Reporter.

The WIKI RoadSMap app allows user to view the information available in the storage system on a map through which it is possible receive all information related to a specific point of the road infrastructure for which there is information on the Web database. The Wiki Reporter app allows the user to locate a point on the map and to provide information relating to that point by sending a ‘Report’ to the Web database storage. The WIKI RoadSMap platform may be available on Internet and on smartphones, both providing different levels of access and subscriptions for those who want access to more specific technical information, aiming to increase road knowledge from different points of view (public, private, commercial, technical, social).

Acknowledgements

This work was partially financed by the PRIN project USR342. The authors wish to thank the Ministry of Education, University and Research, which supported this research, and the Italian Agile company that participated in the project.

References

[1] Carrara, W., San Chan, W., Fischer, S. & van Steenbergen, E., *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources*, European Commission, Directorate General for Communications Networks, Content and Technology, Unit G.3 Value Data Chain, 2015,

[2] ARTIST, http://www.its-artist.rupa.it/ (accessed 1 October 2018)

[3] Ballis, A., *A Transport Information Agent for Accessing the Different ETIS Data Sources*, Project ID: GMA2/52035/2001. FP5-GROWTH - Programme for research technological development and demonstration on *Competitive and Sustainable Growth 1998-2002*, 2005,

[4] Engelbrecht, J., Booysen, M.J., van Rooyen, G.J. & Bruwer, F.J., *A Survey of Smartphone-based Sensing in Vehicles for ITS Applications*, IET Intelligent Transport Systems, **9**(10), pp. 924-935, 2015. DOI: 10.1049/iet-its.2014.0248.

[5] World Wide Web Consortium, http://www.w3.org/standards/ (accessed 1 October 2018)

[6] Cafiso, S. & Di Graziano, A., *Automated In-Vehicle Data Collection and Treatment for Existing Roadway Alignment*, in 'Efficient Transportation and Pavement Systems: Characterization, Mechanisms, Simulation, and Modeling – Proceedings of the 4th International Gulf Conference on Roads', Qatar, pp. 785-797, 2008.

[7] Cafiso. S., Di Graziano, A. & Pappalardo, G. *Safety Inspection and Management Tools for Low-Volume Road Network*, Transportation Research Record: Journal of the Transportation Research Board, (2472), 2015, ISBN 978-0-309-29571-0.

[8] Cafiso, S., Di Graziano, A., D'Agostino, C., Delfino, E. & Pappalardo, G., *A New Perspective in the Road Asset Management with the Use of Advanced Monitoring System and BIM*, MATEC Web of Conferences, Volume 231, 12th International Road Safety Conference GAMBIT 2018, Gdansk; Poland. DOI: 10.1051/matecconf/201823101007.

[9] Odoki, J.B., Di Graziano, A., Akena, R., *A Multi-criteria Methodology for Optimising Road Investments*, Proceedings of the Institution of Civil Engineers: Transport, Volume 168, Issue 1, February 2015, Pages 34-47

[10] Cafiso, S., Di Graziano, A., Goulias, D. & D'Agostino, C., *Distress and Profile Data Analysis for Condition Assessment in Pavement Management Systems*, International Journal of Pavement Research and Technology, **12**(5), pp. 527-536, 2019.

[11] Fedele, R., Praticò, F.G. & Carotenuto, R. & Della Corte, F.G., *Instrumented Infrastructures for Damage Detection and Management*, 5th IEEE International Conference on Models and Technologies for

Intelligent Transportation Systems, MT-ITS 2017 - Proceedings, art. no. 8005729, pp. 526-531, 2017. DOI: 10.1109/MTITS.2017.8005729

[12] Klein, L.A., *Sensor and Data Fusion for Intelligent Transportation Systems*, , Volume: PM305, SPIE Press, 2019. ISBN: 9781510627642

[13] Chodorow, K., *Mongo DB: the Definitive Guide*, O'Reilly Media, Inc., 2013.