LEVERAGING CROWD-SOURCED ROAD DEFECT INFORMATION FOR ROAD QUALITY ASSESSMENT

Dimitrios Efthymiou*, MSc
PhD Candidate, Laboratory of Transportation Engineering
School of Rural and Surveying Engineering
National Technical University of Athens
9, Iroon Polytechniou St., Zografou Campus, 15780
Tel: 0030-210-7721675; Fax:0030-210-7722629
E-mail: defthym@mail.ntua.gr

Constantinos Antoniou
Assistant Professor, Laboratory of Transportation Engineering
School of Rural and Surveying Engineering
National Technical University of Athens
9, Iroon Polytechniou St., Zografou Campus, 15780
Tel: 0030-210-7722783; Fax:0030-210-7722629
E-mail: antoniou@central.ntua.gr

ABSTRACT

A large number of initiatives aims in developing a risk benchmarking procedure for road infrastructure (e.g. iRAP and EuroRAP). Such initiatives can help governments and stakeholders improve the safety of their infrastructure through the optimal allocation of the available resources in the most effective way.

In this paper, an alternative, crowd-sourced approach to data-collection is considered. In such a concept, individuals are invited to report any road defects that they encounter through a convenient web interface, indicating the location, the severity and the type of the defect. This information is aggregated and can be later analyzed. The data used in this research come from the PIN project (http://www.msfree.gr/pin/), an initiative of the Road Safety Institute Panos Mylonas (IOAS).

Key words: Crowd-sourced data, road defects, Pin-Project, Greece

1. INTRODUCTION

Data collection and mapping projects usually fall within the remit of the authorities (mainly because of the high budget). However, during the last few years, new technologies and people's habits have been emerged, contributing to the development of initiatives for the collection of spatial information (e.g. http://openstreetmaps.org). More specifically, 1) the Web 2.0, 2) the social media and 3) the smartphones, compose the technological state-of-the-art and have supported development of these new data-collection methodologies.

Under these new circumstances, the general public has the chance to participate (with an alternative way) in data collection and planning. This intentionally or not participation of the public in data collection and planning is defined as *crowdsourcing*.

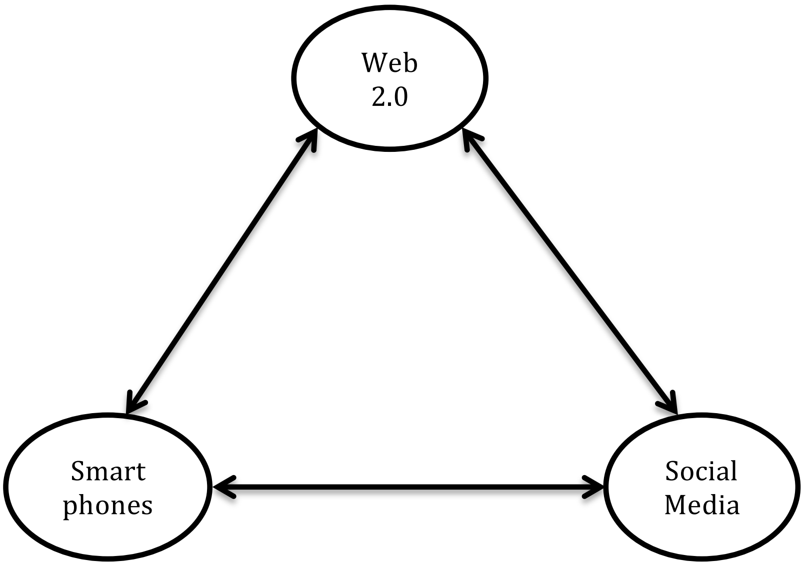

Figure 1. Interaction between Web2, Social Media and Smartphones

With crowdsourcing, volunteers perform the work that used to be made by professionals. Well-known examples of web sites that successfully use the "crowds" to collect data are the free on-line encyclopedia *Wikipedia* (wikipedia.org), the free on-line maps *Open Street Maps* (openstreetmaps.org) but also any kind of Open Source software. Among the advantages of using crowd-sourced data, are that they are: 1) free for use by the public, professionals and researchers, 2) easily, instantly accessible, 3) usually up-to-date, and 4) collected by the latest means of technology. The quality of the resulted dataset differs among the platforms, and depends on the amount of the participants, the administrators and the time of the data-collection process.

The rest of this paper is structured as follows: The term *Volunteered Geographic Information* is defined, and the main crowdsourcing platforms are described in section 2. In section 3, the existing road quality assessment programs (such as iRAP) are analyzed. Finally, the crowdsourcing data collected by volunteering organizations in Greece are qualitatively and quantitatively analyzed, and the results are classified.

2. VOLUNTEERED GEOGRAPHIC INFORMATION

2.1. VGS and VGI

The term *Volunteered Geographic Information* (Goodchild and Glennon, 2010) is used to describe the voluntarily public contribution for geographical data and information collection. The most known example of a platform hosting VGI is the OpenStreetMaps (OSM), which actually is a map of the Earth, freely accessible and editable by the users. The OSM database (composed by the road/rail/waterway network, the images but also by other spatial variables) is widely used by transportation researchers and consultants. The contribution of numerous users in gathering the geographic information, results in systems of quality sometimes higher than those developed by a single user, even if he is specialize (Goodchild and Glennon, 2010).

The term *Volunteered Geographical Systems* (VGS) (Savelyev et al., 2011), is used for the integration of social networks and VGI. Savelyev et al. (2011) give an example of using smartphone applications that integrate VGI and social media for emergency response situations. VGS are based on the willingness of people, not only to report the cases, but also to actively respond to other requests (e.g. www.avego.com)

2.2. Social Media

Social Media are web 2.0 applications that allow their users to share their thoughts on-line. These platforms are addressed to people with specific aims (e.g. www.academia.org for exchange of scientific knowledge, www.linkedin.org for job-finding or recruiting), or to the general public (e.g. plus.google.com, www.twitter.com, www.facebook.com and www.flickr.com). Social media have been used for different purposes: 1) political campaigns, 2) marketing, 3) news updates, 4) recruitment or even for 5) protests revolutions (Efthymiou and Antoniou, 2012).

The multi-variation of the demographic characteristics of their users (different ages, nationality, education level e.t.c) renders these platforms, tools suitable for data collection. Social networks have been used for transportation-related applications (e.g. Efthymiou and Antoniou, 2012; Amey et al., 2011; Bregman, 2011; Grigolon et al., 2011; Carvalho et al., 2010). More recently, Crooks et al. (2012) analyzed spatial geo-referenced tweets, to show how twitter can be used as a distributed sensor system for earthquakes. Moreover, *flickr* and *facebook* include functions for geo-referenced image web uploading, creating new prospects for their usage as tools for road quality information collection.

2.3. Examples of Crowdsourcing Platforms

2.3.1. SeeClickFix

SeeClickFix (www.seeclickfix.com) is an on-line platform where the public can report *neighbor issues and see them fixed*. Its role is intermediary and links the public with the policy makers. Citizens, media organizations, community groups and governments, use SeeClickFix for different purposes. Once a volunteer reports an issue, another can access, vote, edit the description and add information about it. The more the contribution from the users, the more possible it is that the local governments will 'fix' the deficiency. SeeClickFix uses all the characteristics needed to improve life of citizens: Transparency, collaboration, scale, efficiency, simplicity (www.seeclickfix.com).

2.3.2. The Pin-Project

The Pin-Project was launched in 2007 by the Road Safety Institute Panos Mylonas (IOAS) and is hosted on-line at the portal: www.msfree.gr/pin/. The aim of this initiative is to motivate the users of the Greek road network to report voluntarily, by adding a “pin” with (X,Y) coordinates and a short description on a map, any constructional defects (e.g. slippery surface, blind spots, bad lighting, barrier on the street, potholes) that increase the risk of road accidents across the Hellenic motorways. The application is not available in English, which means that its use is restricted only to Greek language speakers.

The first goal of this project was to create a spatial database with the Greek streets defects. Then, the regional volunteering groups undertake to curry out autopsies at the locations indicated in the map, in order to determine the size of the problem, take pictures, and write a more detailed description. Then, a report was written and submitted to the Mayor of each Municipality. Finally, the pin on the map should be updated by changing change shape and color, in order to enable the user to know the stage of the process (however, this step was not finally implemented).

Pin-Project accepted up to 650.000 visitors until May 2009, 12.421 points had been "pinned" and according to the users, 7.468 could be fixed immediately. 1.081 of these points were characterized as dangerous for serious traffic accidents, while the majority of the pins (69%) were about the existence of dangerous potholes on the road surface (Danelli – Mylonas, 2009). The Geographical Information System (GIS) created, offers a valuable spatial database, which has so far been exploited only by the Road Safety Institute.

2.3.3. The Illegal-Signs Project

The web page www.illegalsigns.gov.gr (with the motto: *without illegal signs*) offers a platform where volunteers can report the existence of illegal marketing signs across the Greek street network. The Hellenic Highway Directives indicate that *signs that attract out attention and are placed on roundabouts, side-walks ad points where could cause accidents, perpendicular to the traffic flow, are illegal, dangerous and should be dismantled*. This had also been decided by the «Council of the State». The *Prosecutor of Athens* decided that death caused by an accident on these signs, is a murder by negligence. At least 8 persons per year die in car accidents where an illegal sign was involved, while the one tenth of the car accidents is because of the crash on illegal signs (Illegal Signs Project, 2012).

The signs are first reported by volunteers (who indicate the exact location, attach a photo and make comments), and then are removed by the local authorities (Municipalities, Department of Transport and sometimes the office of Archaeological Athenian sites).

3. ROAD ASSESSMENT PROGRAMS

The most well known Road Assessment Programs are under the umbrella of the non-profit organization named iRAP (i.e. International Road Assessment Program). iRAP cooperates with governmental (e.g. the European Commission) and non-governmental organizations -such as the Foundation for the Automobile and Society (FIA), the World Bank, regional banks and other donors (automobile clubs, individuals e.t.c). The international experience gained from the RAP projects in the developed countries, has been used to create four basic protocols in order to improve the road safety of others. These protocols are the: 1) Risk maps, 2) star ratings, 3) safer roads investment plans and 4) performance tracking.

iRAP Star Ratings are performed either by 1) drive-through or 2) video-based road inspections. The drive-through road inspections require the RAP Inspection Device (RAPID), which is composed by a video camera, a laptop, and a Global Navigation Satellite System (GNSS) antenna. In video-based road inspections, roads are recorded by specially equipped vehicles and then they assessed by «Raters» in a desktop.

A «Road Protection Score» (RPS) is then calculated for each link of 100 meters, using the iRAP software (based on the EuroRAP and AusRAP models). The iRAP RPS has been developed upon the evaluation of risky road infrastructure factors that influence the crashes of car, motorbike or bicycle users and pedestrians. These factors are mainly related with the road environment, such as: the absence of lane strips, the distance of hazardous objects from the roadside, the condition of the road surface and the absence of pedestrian footpath. Each category of factor has assigned a risk value according to its severity. The RPS measures the likelihood of crash occurring in each segment (iRAP, 2011).

Another study in Iowa, used different road assessment methodologies (such as the PLANSAFE models and the usRAP) and concluded that their incorporation in the transportation planning processes can assist the decision makers in monitoring the quality of roads and achieving the goals for road safety (Gkritsa et al., 2011).

4. METHODOLOGY

The the coordinates of the network and geo-referenced road defects were transformed in the Coordinate Reference System (CRS) Greek Grid and network lines were split in links. In order to overcome the issues raised by the quality of the features, such as 1) the different vector type of data (point attribute) and network (linear attribute) and 2) the points were not falling precisely on the line (either because of the users error when placing the pin, or because of the width of the road network), we increased the width at 7 m and "absorbed" the data points falling in that area, in each link. A number referring to the total number of "score" was then assigned to each link. Table 1 presents the basic characteristics of the different types of the network downloaded from the OSM. Then, a linear model based on Ordinary Least Square (OLS) was estimated. The dependent variable was the "total score" of each link, and the observatory were the type of the link (residential, primary, secondary, tertiary, motorway or trunk), the percentage of people with university degree, the population density of the Municipality that they are located in and their length in meters. The results show that the score differs with the type of the road, while the education level and the population density indicates the limitations of this particular way of data collection, since the web users are usually people of high education level (with university degree). Moreover, the higher density of a municipality shows that the denser the area, the more people will report the road defects. Figure 2 shows the road defects and the degree of severity of a region in Attica. Figure 3 describes the steps of these two projects, the OSM, but also the methodology that was followed for this research.

Table 1: Basic characteristics of the OSM network of Attica

Type of road	Length (m)	Links	Total score
Residential	174641	1957	4072
Motorway	35889	115	196
Primary	113407	721	1466
Secondary	100556	809	1612
Tertiary	75458	740	1613
Trunk	12282	60	202

Linear Regression:

$$Y = \beta_{rs}X_{rs} + \beta_{pr}X_{pr} + \beta_{sc}X_{sc} + \beta_{tr}X_{tr} + \beta_{mt}X_{mt} + \beta_{tn}X_{tn} + \beta_{edu}X_{edu} + \beta_{den}X_{den} + \beta_{len}X_{len}$$
(1)

Table 2. Model Estimation

Variable	Description	β	Standard error	t-value
$\boldsymbol{\beta_{rs}}$	Type: Residential	1.21	0.16	7.65
$\boldsymbol{\beta_{mt}}$	Type: Motorway	0.91	0.23	4.02
$\boldsymbol{\beta_{pr}}$	Type: Primary	1.18	0.17	7.00
$\boldsymbol{\beta_{sc}}$	Type: Secondary	1.18	0.16	7.37
$\boldsymbol{\beta_{tr}}$	Type: Tertiary	1.34	0.16	8.28
$\boldsymbol{\beta_{tn}}$	Type: Trunk	2.44	0.28	8.75
$\boldsymbol{\beta_{edu}}$	High education	1.79	0.35	5.17
$\boldsymbol{\beta_{den}}$	Low density (<500 persons/km^2)	1.00E-05	4.00E-06	3.57
$\boldsymbol{\beta_{len}}$	Logarithm of link length	0.10	0.029	3.42

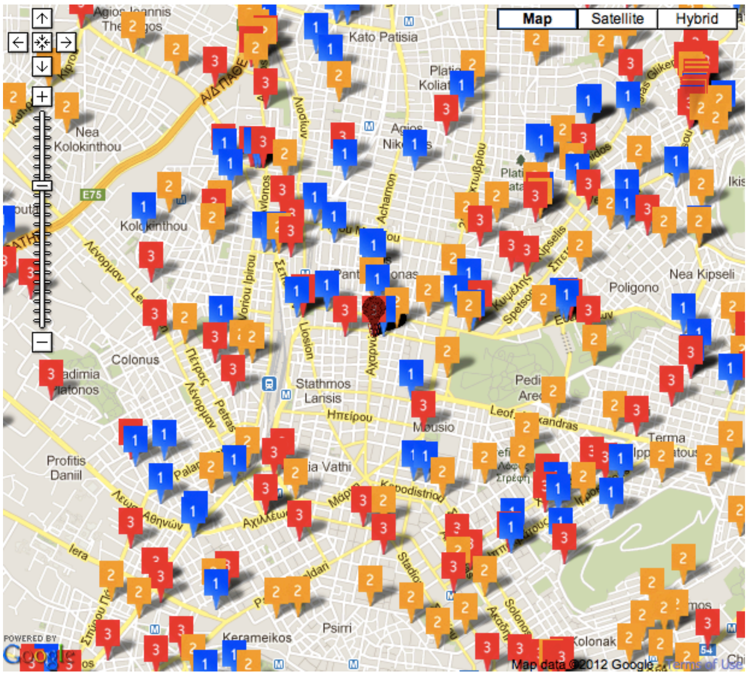

Figure2. Road defects in Attica (color and number show the degree of severity from 1 to 3)

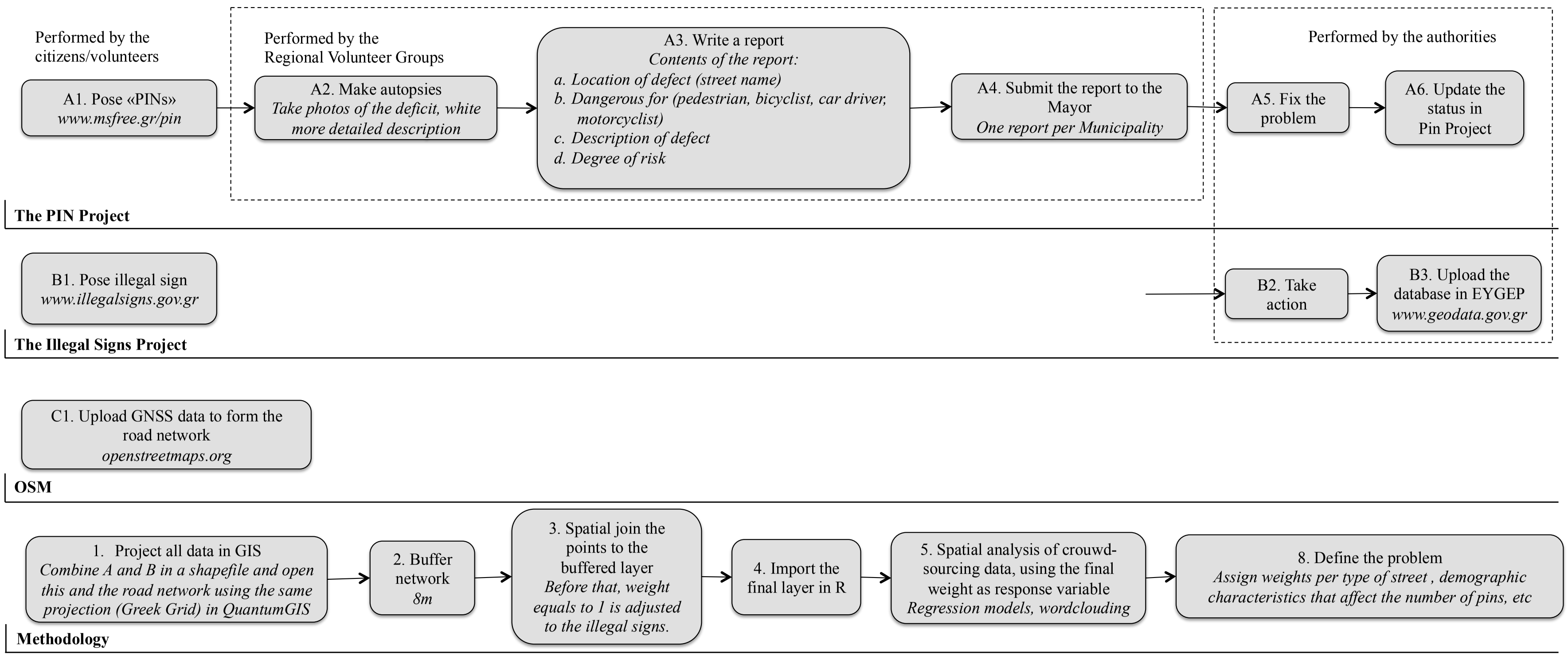

Figure 3. Methodology

5. CONCLUSIONS

The triptych web, social media and smartphones, create a powerful tool for volunteer data collection. Crowd-sourced data compose valuable databases that can be used by the research community. Two widely used platforms for crowd-sourced data collection in Greece are the Pin-Project (operated by the Road Safety Institute Panos Mylonas) and the "Illegal Signs Project". The first aims to motivate the users of the Greek road network to report any structural defects (e.g. slippery surface, blind spots, potholes) that increase the risk of road accidents, and the second to report the existence of illegal signs, which can also be proven dangerous for the drivers if placed perpendicular to the traffic flow, and should be dismantled. Moreover, the road network of the Attica metro area was downloaded from the OpenStreetMaps database (www.openstreetmap.org) and the defect data were linked to it.

The spatial distribution of segments with defects of different type and weight is first presented, and appropriate statistical methodologies are then applied. The resulting index of quality of the road infrastructure is correlated with several possible contributing factors. This risk rating -which can be considered a proxy that indicates the degree of safety for the drivers- is intended to be used as input for further applications. While the available data have severe limitations, they allow for an initial demonstration of the potential use of crowd-sourced road assessment data. The use of such data may be particularly important, especially until a formal road safety assessment program is in place. The results show that the score of network links is a function of constructional and demographic characteristic. People of higher education are more intend to contribute voluntarily for data collection.

Acknowledgements

The authors would like the thank the Road Safety Institute Panos Mylonas for giving their permission of using the Pin-Project's database, but also the administrators of the portal www.msfree.gr/pin who are responsible for the management of these data.

References

Amey, A., J. Attanucci and R. Mishalani (2011). "Real-Time" Ridesharing – The Opportunities and Challenges of Utilizing Mobile Phone Technology to Improve Rideshare Services. Proceedings of the 90th Annual Meeting of the Transportation Research Board, January, Washington, D.C.

Bregman, S. (2011). What's the worst that can happen? How to stop worrying and love social media. Proceedings of the 90th Annual Meeting of the Transportation Research Board, January, Washington, D.C.

Carvalho, S., L. Sarmento and R. Rossetti (2010). Real-time sensing of traffic information in twitter messages. Proceedings of the 4th Workshop on Artificial Transportation Systems and Simulation ATSS at IEEE ITSC 2010, Madeira, Portugal, 19-22 September, 2010

Crooks A. T., Croitoru A., Stefanidis A. and Radzikowski (2012). J.Earthquake: twitter as a Distributed Sensor System. *Trans- actions in GIS.*

Danelli-Mylonas V (2009) .Road Safety: Actions and perspectives. In *Swedish Trade Council Conference,* R.S.I. "Panos Mylonas".

Efthymiou D. and C. Antoniou (2012). Use of Social Media for Transport Data Collection. Transportation Research Arena, Athens

Gkritsa K., Souleyrette R. and Wang T (2011). A Transportation Safety Planning Tool for the City of Ames. Institute for Transportation, Iowa State University

Romm, J. (2006), *Hell and High Water: The Global Warming Solution.* Harper Collins, New York.

Goodchild M. (2007).Citizens as sensors: the world of volunteered geography. *GeoJournal,* Vol. 69, No. 4, pp. 211–221.

Goodchild M. and J. A. Glennon (2010). Crowdsourcing geographical information for disaster response: A research frontier. *International Journal of Digital Earth,* Vol. 3, No. 3, pp. 231–241.

Guo, J. Y. and C. R. Bhat (2007). Population Synthesis for Microsimulating Travel Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014,* Transportation Research Board of the National Academies, Washington, D.C., pp. 92– 101.

Harwood D. W., Bauer K. M., Gilmore D. K., Souleyrette R. and Hans Z. N. (2010).Validation of the usRAP Star Rating Protocol for Applicatio to Safety Management of U.S. Roads. In *Transportation Research Board,* Transportation Research Board.

iRAP/EuroRAP (2011). *Crash rate - Star Rating comparisons. Review of available evidnece.* Working paper 504.2, European Road Assessment Program, EuroRAP.

iRAP (2011). *Safer Roads Investment Plans: The iRAP Methodology.* 504.05, iRAP.

Illegal Signs Project (2012). Website: www.illegalsigns.gov.gr/?page_id=140

R Development Core Team (2011). R: A Language and Environment for Statistical Comput- ing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051- 07-0.

Savelyev A., Janowicz K., Thatcher J., Xu S., Mulligann C. and Luo W (2011). Volunteered geographic services: Developing a linked data driven location-based service.