Sustainable Urban Mobility Boost Smart Toolbox Upgrade

M. Sostaric[1,*], M. Jakovljevic[1], K. Vidovic[2] and O. Lale[1]

[1] Faculty of Transport and Traffic Sciences, Zagreb, Croatia
[2] Ericsson Nikola Tesla d.d., Zagreb, Croatia

Abstract

SUMBooST2 research develops universally applicable data science methodology which extracts key urban mobility parameters and origin/destination matrices from the anonymized big data set gathered from telecom operator. The methodology (toolbox) provides transport planners with a method for fast, efficient, and reliable provision of data on movements within the certain area. Origin/destination matrices with modal split will provide transport planners with valid input data for the planning of urban transport systems. The algorithms which separate relevant mobility data from the overall dataset are the unique part of the toolbox. The algorithms to identify passenger car trips are developed in 2020 project SUMBooST, and they are being upgraded in 2021 to detect trips made by active mobility modes and public transport. For the methodology to be valid, it must be implemented in representative number of cities. Previous SUMBooST project included implementation and validation in the City of Rijeka, and SUMBooST2 continues with two other cities, City of Zagreb, and City of Dubrovnik. The aim of the paper is to present innovative toolbox for the boost of sustainable urban planning based on big data science.

1. State of the Art

Development of ICT has caused that data is generated at an extraordinary scale, leading to their growing amounts. This massive generation of data provides new opportunities for discovering new values. Big Data is the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value [1]. Mobile networks have become both generators and carriers of Big Data, since an enormous amount of data is generated as a result of both user-related and network-related activities [2]. This type of data usually originates from two potential data sources within the mobile network. The first set of data consist of database containing logs of telecommunication activities initiated by the users and is referred to as CDR

(Call Data Record) data. The second set is generated as a result of constant communication between network elements (including terminals) and is referred to as signalling data. CDR represents the data generated by events and represents the records that are created when some telecommunication activities are performed by subscribers in the network. Events include activities such as voice calls, SMS or MMS services, data traffic usage, roaming information, etc. The location information obtained from the mobile network is not as precise as the GNSS location data, but those location uncertainties can be, to some extent, mitigated by a large amount of data. The mobile network data are low-cost alternative to traditional GPS data since the network data can provide anonymous information for millions of users [3]. The mobile network positioning accuracy might vary from 50 to 500 meters with median value of 250 meters in densely populated urban areas [4], and that is considered acceptable for urban

---

*Corresponding author. Email: msostaric@fpz.unizg.hr

mobility use cases. For different positioning methodology used in mobile networks consult literature [5]. There are several cases of application of mobile network data in transportation systems that were subject of various studies. Such cases take advantage of the fact that besides the fact that these data are not captured for transport applications they have significant relevance to their understanding.

Majority of studies have been performed to determine the Origin/Destination Matrix (ODM) from those types of data [6]. Beside origin/destination matrices, multiple studies are dealing with the application of telecom Big Data in public transport planning [7], urban mobility planning [8], [9], transport mode detection [10], urban mobility estimation [11], traffic flow analysis [12] and the reconstruction of human mobility in general [13].

Few papers are dealing with the potential of the integration of Big Data into sustainable mobility planning support systems, and they are on a conceptual level investigating how Big Data can innovate urban mobility policy [14]. In general conclusion is that, while heuristic relevance of Big Data is increasingly investigated, the contribution of this new knowledge source to the policy processes still requires deeper analysis. Besides, the implementation of data-driven planning support system can have multiple implications in favor of improved quality of life and smart and sustainable growth of the future urban agglomerations [15].

The conclusion resulting from the literature review is that there is a significant potential of mobile network data in transportation domain, especially when compared to traditional data acquisition methods.

2. Introduction

The main focus of the research was to address the problem of globally frequent overuse of unsustainable transport modes in urban areas. Transport planners need information on where and how people travel within the city (origin/destination matrix by transport mode). Using this information, they can encourage sustainable mobility modes and dissimulate passenger car usage. This kind of data is complicated, expensive, and time-consuming to obtain by traditional methods.

The defined problem is tackled with the SUMBooST2 project using the innovative method of obtaining the origin/destination matrix correlated with the modal split. Innovative methodology (toolbox) is based on Big Data science further enhanced and validated through traditional traffic research. The most creative and innovative part is the extraction of the mobility parameters from the Big Data set gathered from telecom operators.

SUMBooST2 is a follow-up of SUMBooST 2020 project where the methodology of the toolbox was based on the identification of passenger car trips only. The project further improves the methodology and mainly focuses on identifying sustainable transport modes and their strengthening, but passenger car trips are also analyzed as an additional validation of the first version of toolbox.

The toolbox usage will provide its users with information on origin/destination urban zone pairs with a significant share of trips made by sustainable modes of transport and give them a valid input where they can make upgrades and additionally expand the usage of sustainable mobility modes. Also, the toolbox will provide users with origin/destination pairs with low sustainable transport and a large share of passenger car trips and give them input where to discourage those unsustainable mobility modes.

The main objective was to provide the city leaders and planners with a toolbox that can give them valid input for their transport related decision making. The toolbox provides fast and efficient way to obtain an accurate data set based on which city planners can develop new solutions.

This paper focuses on the description of the new segments of methodology and algorithms for the detection of sustainable modes of transport from the Big Data set. The paper shows a brief overview of the SUMBooST 2020 project which set the basis for the SUMBooST2 project. Afterwards, the preliminary results of SUMBooST2 project are described.

3. SUMBooST

Sustainable Urban Mobility Boost Smart Toolbox (SUMBooST) project has resulted in a proven and validated methodology for fast and efficient transport data collection, fusion, and analytics needed for the transport planning processes. The results show that the methodology and related activities open a new dimension of Big Data usage in transportation engineering, enabling quick, efficient, and safe mobility patterns analytics. SUMBooST aimed to:

- use Big Data analytics and field research to identify pairs of urban zones with a high percentage of passenger car commuters,
- identify reasons for the high number of passenger car trips,
- define and propose measures for the modal shift from passenger car to sustainable transport modes.

The methodology was set up to analyze the 'as is' mobility situation in the coverage area first. The analysis was conducted based on the Big Data research and field research. Big Data research included the process of collecting, organizing, and analyzing large sets of anonymized data gathered from mobile telecom operators to obtain information on daily migration patterns, which are important for urban mobility planning. Big Data research (data science) on anonymized mobile telecom data sets for migrations represents an innovative approach and should be validated.

The results of the Big Data analysis were validated by the results of traditional field research. Field research was performed through an online and phone survey on

commuter patterns, through analysis of traffic flow distribution based on automatic license plate recognition system (ALPR), and through analysis of traffic flow volume and structure (traffic counting). The validation was successful, and it confirmed the correlation between the two sets of results, which resulted in the basis for defining transport issues, challenges, and solutions for the defined challenges.

The methodology was successfully completed in a pilot study, and it resulted in a set of possible solutions for modal shift from passenger cars to sustainable mobility modes. Solutions were proposed for each pair of zones with high share of passenger car trips. The local public and stakeholders confirmed the quality of the proposed transport solutions (proposed measures).

The main result and conclusion are that the whole SUMBooST process was successfully carried out and that the innovative toolbox delivers valid output. The scheme of the proven methodology is shown in Figure 1.

The entire process includes roughly 150 steps. For the illustration purposes, they are visualized in the Figure 2. The main segment, analysis of identified zones with an identification of transport problems is shown in detail in the Figure 3. Besides the definition and the description of activities that must be performed within the SUMBooST toolbox, the process includes a high-level definition (description of the required input data, definition of expected result data, definition on deliverables) and the information on expected validation steps. All activities are logically and sequentially connected, so that, in general, the outcome of the previous activity represents input data for the subsequent activities.
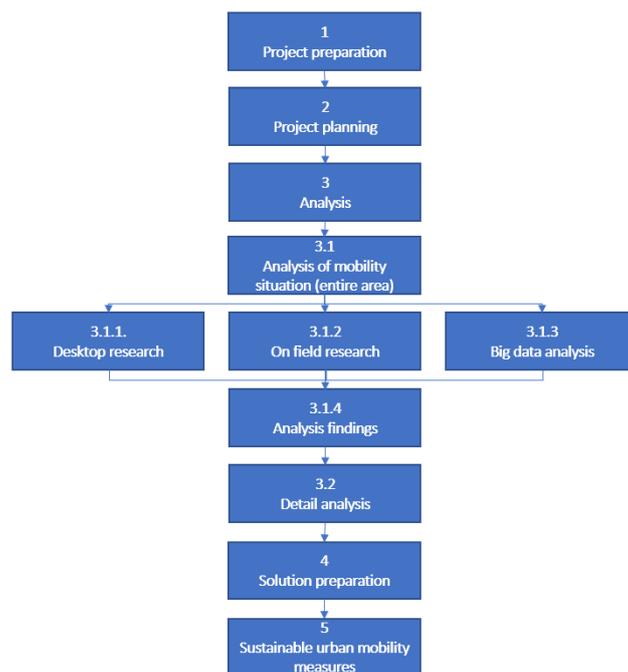

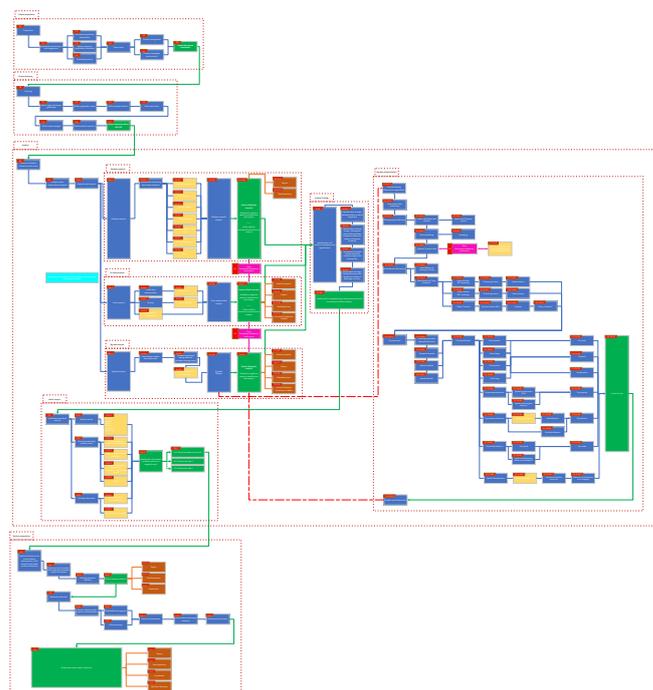**Figure 1.** High level visualisation of SUMBooST toolbox methodology

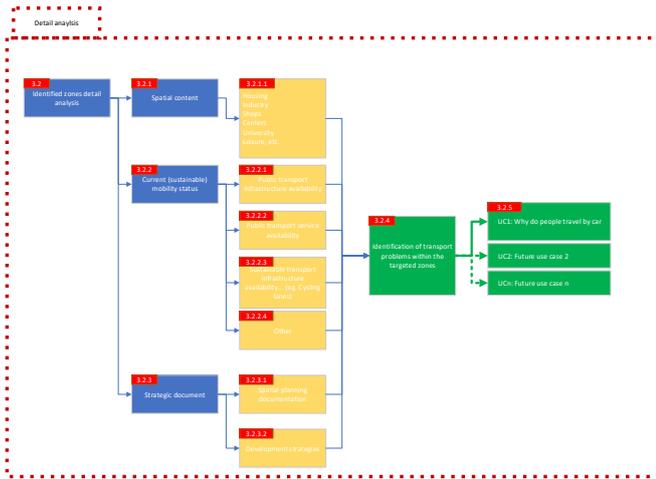**Figure 2.** Illustration of steps within the SUMBooST toolbox

Figure 3. Detail analysis steps

Within the SUMBooST toolbox, data analytics and visualization module has been developed, and system prototype demonstration in operational environment has been performed (TRL 7 reached). The developed web-based tool enables visualization and basis analytical queries on validated Big Data sets. Several data sets are available within the tool, including data sets containing number of migrations (data set for entire characteristic working day, dataset for peak hour, data set for various time frames within the day), and data sets containing information on proportion of migration related to usage of passenger cars and public transport. Besides availability of different types of queries and visualization, toolbox supports data and visualization export for further analysis. Example of toolbox graphical user interface is presented in Figure 4.
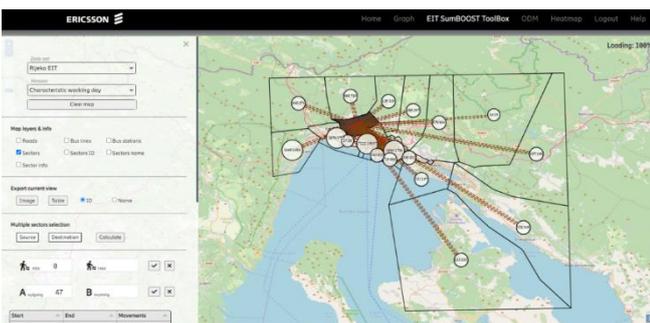

Figure 4. Example of toolbox graphical user interface

To preserve user privacy and to ensure compliance with GDPR directive, several measures have been applied, including space and time clustering (aggregation) and K-anonymity. This guarantees that the individuals who were the subjects of the anonymized data cannot be re-identified while the data remain practically useful. To preserve K-anonymity, all values of migrations lesser than five among sector pairs for any time period are not present on map. For example, if during the analytics it is determined that a total number of identified migrations between sector 24 and 35 during peak hour is '4' or less, this value will not be presented, and the value will be substituted with zero. This impacts less than 1% of all trips identified during the characteristic day.

4. SUMBooST upgrade

Good quality results from 2020 encouraged the project team to further improve the methodology and expand the transport mode coverage with active modes and public transport, which led to a SUMBooST2 project. The project plan was to further improve the 2020 methodology and to mainly be focused on identifying sustainable transport modes and their strengthening, but passenger car trips were analyzed as an additional validation of the first version of toolbox. For the methodology to be valid, it must be implemented in representative number of cities. SUMBooST included implementation and validation in the City of Rijeka as a pilot, and SUMBooST2 continued with the City of Rijeka and two other cities, City of Zagreb as the Croatian capital and City of Dubrovnik as a touristic destination.

During the implementation of SUMBooST project, the COVID-19 pandemic has shown that the entire transport system sometimes needs to adapt and react fast to new requirements that might change rapidly. The project team realized that a SUMBooST toolbox, as a new transport planning tool, can quickly respond to such changes and adjust transport system to new circumstances. The lessons learned during the pandemic will improve the way that mobility is managed in the future. The positive changes will include usage of proposed fast, cost efficient and responsive urban mobility management toolbox based on Big Data sets and data science that is being developed within this project. Lessons learnt during the development of SUMBooST project were included in the development of SUMBooST2 project and SUMBooST2 methodology can support transport system development during such extraordinary situations like world pandemic.

4.1. Methodology

In 2021 SUMBooST2 project, the focus was on upgrading the toolbox with algorithms that can identify trips made by sustainable modes of transport, public transport, cycling or walking and on upgrading the accuracy of passenger car travel identification algorithms.

The first step was to define cities for toolbox validation. The selection of cities was such that it covered as wide range of urban features as possible to prove the universal applicability of the toolbox in as many world cities as possible. The project scope area included three Croatian

cities, Zagreb, Rijeka, and Dubrovnik. Each of the cities has its own specifics that are reflected also in the city's transport system and mobility. These specifics were partly the reason why the cities were selected as the pilots. The project focused on sustainable modes of transport - cycling, walking and public transport and the goal was to find cities in which a specific transport mode is presented with significant demand. The City of Zagreb is the biggest city in Croatia and cycling is the most accepted as commuting mode among other Croatian cities. Therefore, it was relevant for the analysis of cycling flows and piloting the data science algorithms for cyclists. The City of Rijeka has a well-defined public transport network with high demand which made them relevant for analysis and testing public transport data science algorithms. The City of Dubrovnik has characteristic walking routes with many pedestrians during the summer months, and specific road transport network which from the project aspect, made the city ideal for researching pedestrian flows, passenger car traffic flows and for testing applicability of data science to pedestrians and passenger cars.

After the definition of the pilot cities, the methodology steps were defined. The research within the project consisted of three main sections, desktop research, field research and Big Data collecting.

After the initial research conduction and input data collection, the next step was the Big Data science. The implementation of the Big Data science should result in the algorithms for transport mode which must be additionally validated with field research results. The final step of research was the creation of an origin/destination matrices with a modal split.

4.2. Research plan

The first step in the methodology was the desktop research for thorough analysis of the traffic system of pilot cities. In addition to the traffic system, a detailed analysis was conducted based on demographics, economic activity, and facilities. The purpose of desktop research was to get to know each segment of pilot cities so that field research and analysis of the transport system can be adequately organized and conducted.

Field research was the next step in which a traffic counting, and surveying were conducted to get the valid data on current state of the traffic system. Each of the three pilot cities was analyzed. Cities of Rijeka and Zagreb were analyzed in a period before summer holidays to get the most relevant transport conditions, and the City of Dubrovnik was analyzed during summer season because that is the period when the most transport issues occur. The field research has resulted in a set of data on characteristic trips by modes of transport and traffic volumes on specific city locations that were used for validation of mobility data extracted from Big Data.

The main purpose of the field research was to build basis for the Big Data validation. During the same period as the field research, Big Data collection was conducted from a

mobile network operator with significant market share. Data was pre-processed and anonymized. The final step within the Big Data project section was the application of the Big Data science to extract the mobility data from Big Data sets.

Big Data validation was the key project activity and the point that defined the success of the project by comparing the traffic parameters obtained from Big Data and from field research. Results of the analysis from those two independent sources must correspond, which proves that the correct mobility data was obtained by Big Data science.

The next step was the identification of origin and destination zones of migrations based on the extracted mobility data. The goal was to find zone pairs with significant usage of sustainable transport modes (for additional development) and with significant passenger car usage (for substitution with sustainable modes).

The goal was to get an answer on questions 'How to improve and perfect sustainable mobility options?' and 'Why people travel mostly by passenger cars?'.

4.3. Research results

Based on the demographic and economic indicators, the urban planning and transport planning documentation, the helicopter view of the transport system of pilot cities was prepared. This step was necessary to present the project team's understanding of the overall transport situation in the cities of Zagreb, Dubrovnik, and Rijeka.

Field research was conducted to calibrate and validate new Big Data methodology on trips and travel habits. Research for certain modes of transport was conducted. In the City of Zagreb, field research emphasized walking and cycling, in the City of Rijeka emphasis was on urban public transport and in the City of Dubrovnik on urban public transport, traffic flow distribution, and walking. In all cities a web survey on commuting habits was conducted.

Big Data analytics included the process of collecting, organizing, and analyzing Big Data sets to discover useful information that can be used by urban mobility stakeholders. This project utilized anonymized Big Data sets originating from mobile telecommunication operator that comply with GDPR regulations.

The key result of this research was the set of data needed to carry out further project activities. For the project to be successfully completed, a key prerequisite was the collection of a specific set of data. That prerequisite was accomplished. Big Data was collected and was ready for the application of Big Data science, and field research data was collected and used for validation of data science results. The data collection process was carried out according to plan in all three pilot cities.

5. Big data science

The mobility data from Big Data set was validated with field research results. The traffic volume from Big Data

was validated by field traffic counting and origin/destination matrix was validated by citizen survey on commuting. Regarding the identification of trips, Big Data science was applied to all modes of transport and each mode has its own algorithm which needed to be validated with field research results. The basic algorithm for identification of passenger car trips was developed in 2020 project so it was slightly upgraded and used in 2021 project with the same validation process.

5.1. Validation

Big Data validation was the key project activity and the point that defined the success of the project by comparing traffic parameters obtained from Big Data and from field research. Results of the analysis from those two independent sources had to correspond, which proved that the correct mobility data was obtained by the Big Data science. The mobility data from Big Data set was validated with field research results such as traffic counting, public transport passengers counting, traffic flow analysis based on ALPR and web household surveys. Some historical data of transport operators were also used as additional dataset for the validation. The results for City of Zagreb were additionally validated through results of the transport model developed in a Master plan of the transport system of the City of Zagreb, and surrounding area. The Faculty of Transport and Traffic Sciences participated in the Masterplan preparation and is familiar with the model. For model preparation a massive field research was conducted. The main focus of field research were household surveys, screen line surveys, public transport passengers counting, vehicle counting, cyclists counting etc. The model was calibrated with those surveys and very high GEH was achieved. Therefore, the project team decided to use this model as an additional source for Big Data methodology validation. The transport model was used to validate a modal split and transport matrices. An additional control of modal split was carried out with the results of web survey.

Regarding the identification of trips, the Big Data science was applied to all modes of transport and each mode had its own algorithm which needed to be validated with conventional data sources. The basic algorithm for identification of passenger car trips was developed in 2020 project so it was upgraded and used in 2021 project with the same validation process. In this project the algorithms for transport mode detection have been upgraded for all transport modes (heuristic rules), and in particular for detection of public transport trips (statistical modelling). The positive validation of the algorithms confirmed validity of the Big Data science which means that all the mobility data extracted from Big Data is useable and relevant for transport planning.

5.2. Preliminary results

The project is ongoing and project team is still working on validation process, but the authors of the paper can give the preliminary results to show the concept of Big Data sets validation.

A good, simple but strong example of validation result are the data sets on traffic flow intensity and traffic flow unevenness in the City of Dubrovnik. For the validation the project team compared the data collected by traditional method (traffic counting with automatic traffic counter) and the data extracted by SUMBooST2 algorithm from telecom Big Data set.

The analysis showed a statistically significant correlation between the hourly load according to the data obtained from Big Data and data obtained by automatic traffic counting ($R2=91.89\%$). Results of comparative as well as correlation analysis are shown in Figure 5 and Figure **6**.
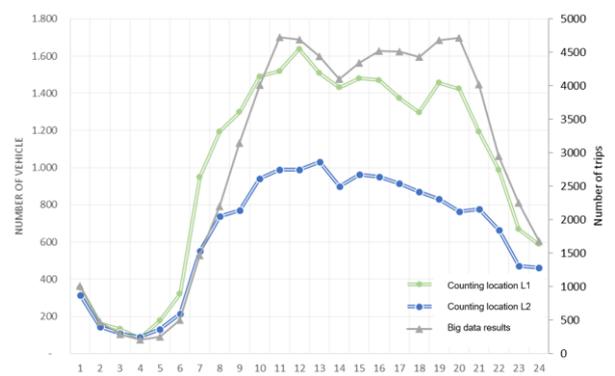

Figure 5. Comparative analysis - daily unevenness of traffic flow (a characteristic day of the week)
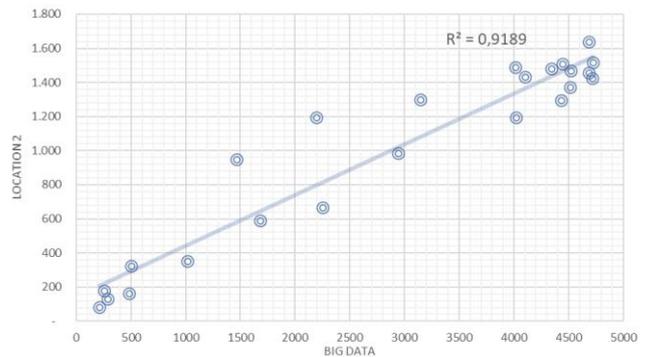

Figure 6. Correlation analysis - daily unevenness of traffic flow (a characteristic day of the week)

6. Next steps and conclusion

The main result of the conducted research and implemented Big Data science is that the conducted methodology (toolbox) will result in a reliable origin/destination matrix with modal distribution.

The next step is to finalize the validation process and analyze the obtained data set and determine the negative and positive pilots of zone pairs of zones (starting zone -

destination zone). Negative pairs of zones are those with a high share of personal car travel. Such zone pairs will be further analyzed to determine the reason for the unfavorable modal distribution and to find a way to encourage the use of sustainable modes of transport. Positive pilots of zone pairs will serve as an example and a guide to a transport system dominated by sustainable modes of transport.

The main conclusion is that the SUMBooST2 toolbox will bring a completely new product to traffic planning processes, significantly simplify traffic system planning and easily identify system needs. Fast and efficient, and most importantly reliable collection of origin/destination matrices with modal distribution is an innovation that every city in the world wants.

Acknowledgements

References

[1] De Mauro A., Greco M., Grimaldi M. A formal definition of Big Data based on its essential features. Libr. Rev. 2016; 65(3):122-135.

[2] He Y., Yu F.R., Zhao N., Yin H., Yao H., Qiu R.C. Big Data analytics in mobile cellular networks. IEEE Access. 2016; 4:1985-1996.

[3] Rojas M.B., Sadeghvaziri E., Jin X. Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. Transp. Res. Rec. 2016; 2563(1):71-79.

[4] Zandbergen P.A. A Comparison of Assisted GPS, WiFi and Cellular Positioning. Trans. GIS. 2009; 13(1):5-26.

[5] Campos R.S. Evolution of positioning techniques in cellular networks, from 2G to 4G. Wirel. Commun. Mob. Comput. 2017; 2017(2):1-17.

[6] Bonnel P., Hombourger E., Olteanu-Raimond A.M., Smoreda Z. Passive mobile phone dataset to construct origin-destination matrix: Potentials and limitations. Transp. Res. Procedia. 2015; 11:381-398.

[7] Zannat K.E., Choudhury C.F. Emerging Big Data sources for public transport planning: A systematic review on current state of the art and future research directions. J. Ind. Inst. Sci. 2019; 99(4):601-619.

[8] Lee S. The use of mobile phone data in transport planning. Int. J. Technol. Policy Manag. 2020; 20(1):54-69.

[9] Milne D., Watling D. Big data and understanding change in the context of planning transport systems. J. Transp. Georg. 2019; 76:235-244

[10] Bachir D., Khodabandelou G., Gauthier V., El Yacoubi M., Puchinger J. Inferring dynamic origin-destination flows by transport mode using mobile phone data. Transp. Res. Part C Emerg. Technol. 2019; 101:254-275

[11] Vidović K., Šoštarić M., Mandžuka S., Kos G. Model for estimating urban mobility based on the records of user activities in public mobile networks. Sustain. 2020; 12(3):838.

[12] Galloni A., Horváth B., Horváth T. Real-time monitoring of Hungarian highway traffic from cell phone network data. ITAT 2018 Proceed. 2018; 2203:108-115.

[13] Simini F., González M.C., Maritan A., Barabási A.L. A universal model for mobility and migration patterns. Nature. 2012; 484(7392):96-100.

[14] Pucci A., Vecchio G., Concilio G. Big data and urban mobility: a policy making perspective. 15th World Conference on Transport Research; 26-31 May 2019; Mumbai, India. WCTR 2019; 2019.

[15] Semanjski I., Bellens R., Gautama S., Witlox F. Integrating big data into a sustainable mobility policy 2.0 planning support system. Sustain. 2016; 8(11):1-19.