

Anna Urbanek\*

## BIG DATA - A CHALLENGE FOR URBAN TRANSPORT MANAGERS

*Big data is a term used to describe large, variable and very diverse data sets, which are very difficult to process and analyse, but on the other hand which are extremely valuable from the point of view of gaining knowledge, for example, on consumer behaviour or real-time changes in a specific market. Automated fare technologies in public urban transport are implemented ever more widely in many cities worldwide. Because of these technologies the public transport organisers have access to large data sets, which create new possibilities in the decision-making process. The aim of the paper is to discuss the role of big data in the decision-making process in urban public transport and also the main barriers to big data use in the urban mobility management.*

**Keywords:** Big data, urban transport, smart card, electronic ticketing, fare collection, public transport, transportation demand.

### 1. Introduction

Information and Communication Technologies (ICT) and geolocation technologies are more and more universally used in every field of life and increasingly strongly affect all its aspects. Investments in modern electronic fare collection systems for the urban public transport services are ever-more popular in cities worldwide. Such systems gradually replace traditional fare collection systems based on paper tickets, the use of which is related to numerous limitations linked primarily with the possibilities of differentiating fares, the need to ensure a distribution network of appropriate reach, as well as with possibilities of acquiring knowledge about the demand for services.

The first electronic tickets originated with the development of magnetic strip card technology in the 60-ies of the 20th century in the United Kingdom. That technology was relatively simple and cheap. The main drawback of cards and tickets with a magnetic strip was the fact that they required a physical swipe and did not ensure a sufficient level of safety, because it was easy to read and reprogram them. Moreover, this technology did not enable the collection of data on the demand for urban public transport services [1]. Modern electronic payment systems for urban public transport based on contactless electronic cards and mobile telephony are relatively new technologies. Such technologies have been implemented on a wide scale only for around twenty years.

Hence the potential to use all data sets generated by these systems in the process of urban public transport management is the topic, which has not been studied enough in the literature of the subject.

### 2. Big data definition

Big data is a term, which – in the meaning used now – appeared in the scientific literature only at the end of 90-ies of the 20<sup>th</sup> century [2]. One of most comprehensive definitions of the big data term was created by Einav and Levin in 2013. In accordance with this definition, big data are big datasets available on a large, massive scale (millions or billions of observations), available in real time or close to it, having many variables, very diversified in terms of kind and type, and also much less structured than the datasets used so far [3]. In 2001 the Gartner company (formerly META Group) created a big data definition in so-called 3V model, which emphasises three main features of so big datasets [4]:

- high-volume (large amount of data),
- high-velocity (high speed/variability of received and sent data),
- and high-variety (high variety of data and sources of their origin).

In addition, in their definition the Gartner analysts indicate that these are the data that require new, innovative

\* Anna Urbanek

Department of Transport, Faculty of Economics, University of Economics in Katowice, Katowice, Poland  
E-mail: anna.urbanek@ue.katowice.pl

forms of processing to support the decision making, to explore new phenomena, and also to optimise and automate processes [4].

One could say most generally that these are sets of information about the surrounding world, to which we have access due to the fact that it is generated by all the equipment featuring an indirect or direct capability to gather, process or exchange data through computer networks. This means in practice e.g. all cameras, sensors, mobile phones, smart phones, cars, objects which could be geolocated, payment cards, RFID readers etc. In the digitised world big data are generated without human interference, beyond intentions and frequently also beyond human awareness, creating a kind of Digital Universe. Big data is not information as such; to become information the data must be appropriately categorised and analysed, which turns out to be not an easy task. Figure 1 presents the forecast growth of data generated in individual years between 2009 and 2020 measured in exabytes (1 EB is approx.  $10^9$  GB). It is estimated that in 2020 the mankind will generate 40,000 EB, of which only 33% could be categorised, analysed and used as information significant from the interesting processes point of view [5].

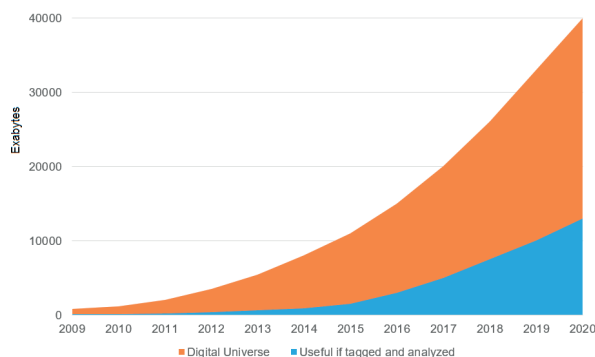


Fig. 1 Data growth between 2009 and 2020.

Source: Own study based on [5]

The big data processing and analysing is difficult, but very valuable. Already today many companies build their competitive advantage on the market exactly due to the data analysis. For example, cookies gathered in an internet browser are used for marketing purposes, to match the offer with individual consumer needs, and special computer software analyses behaviour of people visiting websites so as to optimise planning inventories of appropriate products included in the offer [6].

### 3. Big data sources in urban public transport

With increasingly popular implementation of electronic payment systems for public transport services in cities, the big data analysis and use plays an increasingly great role in the process of urban public transport management. E-ticket systems compared with distribution systems based on paper tickets create huge possibilities in the field of price differentiation as well as of collecting the data on the demand for transport services. Two types of electronic fare collection systems in urban public transport are now being implemented and developed throughout the world, i.e. contactless smart cards and mobile ticketing. Three main types of e-tickets may be distinguished in the mobile ticketing technology [7]:

- premium SMS based transactional payments – user pays fare with the next phone bill or pays for the travel from funds available on his/her pre-paid card,
- OCR (optical character recognition), when passenger/user receives a special code, for example a QR code that contains all needed information,
- NFC (Near Field Communication) technology, where the process is very similar to the OCR technology, but in the case of NFC the information is saved in the NFC memory of the phone.

OCR and NFC technologies become more and more popular, slowly replacing payments via SMS messages. This is related to a widespread use of smart phones and of special applications available for such devices. By means of them passengers have a possibility to buy the ticket of selected public transport organiser and to pay for it without cash by means of a payment card, which the user assigns to his/her individual account. So a smart phone can be not only a carrier of a season ticket, but also an e-money carrier, and may be used as a journey planner as well.

In the case of contactless smart cards technology passengers use a plastic card with an embedded chip, storing the most important information. Such cards need only to place them close to the reader at a distance of approx. 10cm, and they communicate via high-frequency waves like a Radio Frequency Identifier (RFID) [8]. Nowadays this is the most popular technology of electronic fare collection in the urban public transport. The first smart cards started to be used on a large scale in the 1980s. Today they are widely used in banking, health care, government and transportation. The microprocessors used now in contactless smart cards are produced based on EMV (the acronym for Europay, MasterCard and Visa) technology standards. It provides great possibilities in the field of information processing and transaction management, which is very important in the case of dynamic cost and travel time settlement in urban public transport [9]. The functionality of contactless smart cards is very broad:

- they can be e-money carriers and operate as payment cards,
- smart cards can be carriers not only of one, but also of many public transport organiser tickets [10],
- individual concession rights can be encoded on smart cards, even for several public transport organisers,
- smart cards can be anonymous or personalised, i.e. assigned to a specific user, each card has its unique number,
- smart cards can be carriers of single travel and season tickets as well as can operate in dynamic travel cost settlement systems, so-called pay as you go, that is requiring to register the entry to and the exit from the vehicle to calculate the payment for the really travelled distance, time period, number of stops etc. [9].

Substantial part of e-ticket systems operating now worldwide in large cities and metropolises is based on the contactless smart cards technology and on the registration of entering and leaving the vehicle, so-called Check-in/Check-out (it requires that passengers physically register the vehicle entering and leaving by placing a smart card or mobile phone in front of a reader. The system calculates then the due fare and charges it to the passenger account [7 and 11]. The e-tickets technology allows to identify and record all transactions made by passengers, i.e. the place and time of buying a specific ticket type, entering and leaving the vehicle etc. However, attention should be drawn to the fact that the data source in such systems comprises not only smart cards, but also urban transport vehicles equipped with GPS (Global

Positioning System) as well as on-board computers and other devices being an inseparable part of the ticket distribution and control system.

To illustrate the data sources and sizes of data sets obtained by a public transport organiser we can use the example of Silesian Public Services Card (SKUP) project, implemented in 2015 by the Municipal Transport Union of the Upper-Silesian Industrial District (KZK GOP), which is the largest in Poland and one of largest public transport organisers in Europe. Table 1 specifies the most important figures describing the KZK GOP and the SKUP project infrastructure.

The Union is made now of 29 municipalities of the central part of the Silesian Voivodship. The project is also participated by two municipalities (Tychy and Jaworzno), which are not members of the KZK GOP. The project is aimed at establishing a supra-local IT system, which will increase the scope and accessibility of services provided by public institutions via electronic channels and at the same time it will become a tool supporting the process of management in public administration. SKUP contactless smart card is an e-money carrier, which enables making payments not only for the public transport, but also for municipal administrative services, cultural and recreation-collective services, libraries and for parking. So the SKUP card is not only a carrier of season e-tickets in urban public transport, but also allows dynamic settlement of travel cost and time in pay as you go systems in various tariff systems [12].

Selected data related to the KZK GOP and the Silesian Public Services Card project

Table 1

1	Population of KZK GOP municipalities	1,951,603
2	Number of lines (bus and tram)	346
3	Number of vehicles equipped with on-board computers and operating in the SKUP network	1,350
4	Number of mobile modules to collect fares/recharge cards (external sales network and municipal points)	1,210
5	Number of parking meters operating in the SKUP network	223
6	Number of ticket machines	154
7	Number of passenger service points (KZK GOP and external points)	47
8	Number of devices to check the passenger rights in vehicles (so-called 'ticket inspectors')	320
9	Number of SKUP cards issued (as of 31 August 2016)	171,761 (of which 160,359 personalised cards)
10	Number of SKUP cards issued (forecast as at 2017 end)	250,000
11	Monthly number of registered cases of entering and leaving the vehicle in the system (data for August 2016)	4,746,292 (check-ins) 3,439,379 (check-outs)
12	Average monthly number of registered cases of entering and leaving a vehicle in the system (forecast for 2017)	approx. 22,000,000 check-ins approx. 18,000,000 check-outs
13	Average monthly number of sold single travel tickets (paper tickets) in 2015	4,320,975
14	Average monthly number of sold season tickets (paper and electronic tickets for the SKUP card) in 2015	114,331
15	Average monthly number of single travel tickets sold in 2015 via mobile phones and mobile applications	37,827
16	Number of visits on the KZK GOP website (time tables) including via mobile devices in 2015	25,795,862
17	Number of time table changes in 2015	387

Source: Own study based on the KZK GOP figures

The SKUP project was implemented in October 2015, hence it is still not possible to refer to a full system operation. The first year of system operation is a transition period, during which passengers just learn the new rules of tickets purchase and validation and also the process of SKUP cards issuing continues. Hence the table presents the KZK GOP forecasts related to sizes of generated data, assuming a full operation of the system. As Table 1 shows the commissioning of electronic payment system for urban services, of which primarily for urban public transport, requires substantial investments in the IT infrastructure. The data generated by such system users and all its component devices are diversified and create huge sets that require categorisation and analysis.

#### 4. The big data role in the process of public transport management in cities

Automatic fare collection systems are implemented more and more frequently in cities worldwide; they are a source of big amount of data about movements, transactions made and transport behaviour of city residents. This is of particular importance especially in the context of growing population of the cities. The urbanisation progressing in recent years is one of most important processes of economic, spatial, political and social importance. More than a half of the global population (54%) lives now in cities and only 200 years ago the city population made only 3% of the total. In the European Union countries approx. 75% of population live in cities, and it is forecast to grow to around 84% by 2050 [13].

The increasing population of cities and at the same time the expectation of good living and travelling conditions in cities create a challenge for the urban transport, because transport is one of factors enabling the city development. A quantitative development of transport in cities is limited not only by the urban space, the transport network capacity, but also by possibilities to finance investments from public funds. Hence the role of information about the demand for transport services is so great, as it becomes the basis to make decisions in the field of pursuing a more effective use of the possessed resources.

The majority of demand study methods used now, starting from preference studies up to advanced traffic modelling systems, are based on data acquired from entities, occupancy measurements and questionnaire surveys (most often carried out by interviewers), acc. to the rules developed decades ago, when the ICT were not so widely used. Today the data originating from daily events registered by IT systems (generated by the users and acquired from the equipment) can replace many methods used so far and can be used for the needs to build traffic models. The development of a traffic model for a specific area requires first of all the knowledge

about the traffic size on individual routes in specified time periods, the knowledge of traffic sources and destinations, and also about the used means of transport. Electronic payment systems for public transport services provide numerous precise observations on the transport behaviour, which could be referred to time and specific place in the transport system and analysed dynamically in a selected period of time. Traffic models created for the needs to manage urban public transport may be used primarily for the analysis of the existing situation, which is applicable in operational short-term management of the transport offer, and for forecasts, i.e. for long-time planning.

The current management of the transport offer consists mainly in adapting the offer to current needs and expectations of residents. Traffic models allow to carry out simulations and assessments of changes introduced in the transport system and based on that to introduce current changes in the public transport offer. Most often such changes primarily consist in adjusting the vehicles running frequency, modifying the line routes or changing the capacity of vehicles servicing the transport lines. The number of carried out modifications depends on the size and nature of the area, in which the public transport is managed. For example, in the KZK GOP in Katowice approx. 400 modifications of time tables are made every year, consisting mainly of line route changes, changes resulting from including or excluding the service of new stops or changes of the time and frequency of vehicles. Precise knowledge about the number of passengers using urban public transport on a given route during a specific time of day is indispensable for a public transport organiser, because each additional journey of a vehicle on the line means additional costs for the municipality equal to EUR a few dozen or a few hundred thousand annually. None of occupancy calculation methods used so far can provide this knowledge [14].

An equally important application of traffic models consists in long-term, frequently in a perspective of a dozen or so years or several decades, forecasting of changes in the city transport system operation and related changes of transport behaviour, which result from changes in the land development, including implementation of new investments in the city, e.g. new plants, education facilities, commercial and service centres, housing estates and recreation places as well as of other facilities being the traffic sources and destinations [14].

The use of data related to the demand for services, whose sources are the modern ICT technologies implemented in urban public transport, provides benefits to public transport organisers and to passengers as well. The knowledge about passenger flows allows a more effective management of the transport offer (e.g. a better adaptation of vehicles in terms of capacity, optimisation of the number and frequency of journeys), which can result in cutting the expenditures on the operations. A better understanding of transport behaviour

allows primarily to adapt the transport offer to the passenger needs and to increase the public transport attractiveness. Precise knowledge of transactions carried out, that is the tickets type, number and time of purchase, allows a more flexible and innovative pricing policy [9]. Longer-term image improvement can result in increased demand for services and increased revenue from ticket sales. Moreover, increasing the competitiveness and attractiveness of public transport, in particular against individual means of transport, may also be the source of external benefits perceived by all residents of specific area or city. The reduction of travels carried out by the individual transport results first of all in the reduction of congestion and of adverse environmental impact (lower emission of pollutants and noise) and improvement to the quality of life in cities and urbanised areas.

Precise data on passenger flows and ticket revenues on individual lines may be used also to make more detailed mutual settlements between operators or municipalities, who organise together the public transport in a specific area. The scope of data used in such case depends then on the adopted model of financing. In the case of the SKUP project implemented by the KZK GOP, being a union of 29 municipalities, the introduction of rules of settlement with municipalities was one of the main assumptions made for the system. A new model of KZK GOP urban public transport financing assumes using the data from registration of entering and leaving vehicles for the needs to calculate the line profitability. To calculate the ticket revenue for a specific line it is necessary to link the information about ticket (season and single travel) prices with the total number of travelled kilometres during ticket validity periods broken down to lines and municipality areas [15].

## 5. Barriers to big data use in urban public transport

The use of big data generated by electronic payment systems in urban public transport is related primarily to concerns for personal data protection. This is an issue relatively broadly considered in research studies [16]. Each mobile phone, smart phone and each smart card has its unique number. In addition, each mobile phone, smart phone and personalised smart card is assigned to an individual user together with his/her detailed personal data, which are stored in personal databases of mobile network operators, card issuers etc. So the concerns for the privacy protection are common to all technologies, which link a device or card with a specific user and which allow to gather data on his/her behaviour (mobile telephony, bank systems, i.e. credit and payment cards or smart cards used in the health service, urban public transport etc.).

Modern e-payment systems in the public transport based on smart cards satisfy the strictest safety standards, determined

by bank systems. In very many cases these are just banks who are partners of such projects in cities, being responsible for issuing the cards and also for storing and processing the personal data of users. It should be also considered that the issue of personal data protection is regulated in the European legislation in a very restrictive way, in particular in the context of dynamic development of ICT tools, which enable the data gathering and processing on a scale unprecedented so far.

To carry out studies on the transport behaviour and on the demand, a public transport organises needs, apart from the data about mobility and passenger flows, also the knowledge about passengers' age structure and related rights to concessionary travels. In numerous models, in particular those in which the data from e-ticket systems are used for mutual settlements between operators or municipalities organising transport together, there is also a need to gain the knowledge of the municipality of residence [16]. Because of restrictive legislation in this field, the databases containing the personal data of system users are separate from databases used in the process of analysis. That means that in the process of studying the passengers' mobility a public transport organiser should not combine those databases and explicitly identify passengers. Exactly the same processes may be encountered in the bank sector or at mobile networks operators, who for the needs to carry out business and marketing research also analyse the user transactions.

It is also worth noticing that in the literature of the subject more and more attention is paid also to studies on mobility in cities using the geolocation data of residents' smart phones and mobile phones [17]. The generality degree of data acquired from sensors and satellite systems makes them now much less useful than from systems based on smart cards, and the use of more detailed information, e.g. from mobile telephony operators, is very problematic, mainly due to issues related to the personal data protection.

The barriers to the big data use in the process of public transport management should undoubtedly include the cost of an electronic fare collection system implementation. The big data collection most frequently means a systematic or stepwise expansion of IT in the entity. Most often this is an effect of substantial investments in the IT system and it is related to significant funds spent on the system maintenance. This is a costly and time-consuming investment, which requires:

- building an extensive, technologically advanced IT infrastructure (e.g. the issue of building or renting large data processing centres and the necessity to ensure their operation continuity),
- training people, acquiring extensive and specialised knowledge,
- numerous organisational changes.

For example, the total net cost of the SKUP project implementation and its maintenance during 5 years from



commissioning amounts to approx. PLN 190 million (i.e. around EUR 45 million). The process of project preparation and its implementation took approx. 7 years. It should be added that the project was co-financed from the European Regional Fund under Regional Operational Programme of Silesian Voivodship for the years of 2007 - 2013 [12]. Figure 2 presents the main groups of barriers to the big data use in the urban public transport.

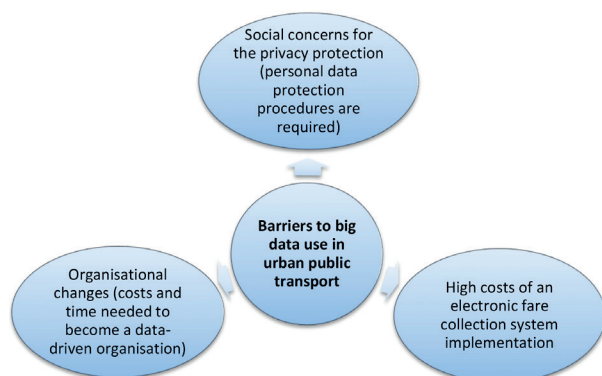


Fig. 2 Main barriers to big data use in urban public transport.

Source: Own study

Moreover, it is necessary to draw attention to the fact that the traffic models development, apart from buying specialised software for urban traffic modelling and forecasting, which is also a costly investment for a public transport organiser, requires structured information. Huge amounts of data received from the system require pre-categorisation and analysis, so that later they could feed not only traffic models, but be useful in the decision making. That requires the next IT tools, specialised knowledge, and also new analytical methods.

## 6. Conclusions

ICT technologies are increasingly widely used in urban public transport, primarily in electronic payment

systems, which are ever-more universally implemented in cities worldwide. Such systems generate huge and diversified datasets, providing numerous precise observations on the transport behaviour and passengers mobility. This data can be the source both of current information about the market, available almost in real time, and also can become the basis for long-term analyses and plans. None of hitherto known and used research methods can provide such knowledge. The progressing urbanisation and the striving for an increase in the effectiveness of provided services make that the role of using the information brought by big data in the process of mobility management in cities has been growing.

The possibility of collecting accurate data about the demand for services, which beyond any doubt is the greatest advantage of those systems, on the other hand is a significant element raising some social concerns. They refer first of all to maintaining the privacy of users and safety of personal data. Such concerns frequently result also in a very slow process of such systems acceptance in society [18].

The performed analysis shows that the role of data acquired from the ICT systems has been permanently growing in the process of collective public transport management. This determines a new direction of changes for this sector, in which the entities managing the urban public transport to a larger and larger extent will transform into so-called data-driven organisations. In the past the data acquisition was the main problem of mobility and transport behaviour analyses. Today the vastness and diversification of received data make that the problem consists in their categorisation and analysis, so as to make them useful in decision making. The analysis of the subject literature and of research performed in this field shows that the possibilities to use the data in the process of urban public transport management are only being studied. This is the area of utilising the scientific knowledge in the field of economy, transport engineering, and information science as well as of cooperation between the sector of science and the business and the transport policy entities on various levels.

## References

- [1] TURNER, P., SMITH, B.: *Integrated Electronic Ticketing - Field of Dreams?* Eighth World Congress on Intelligent Transport Systems, Sydney, October 2001
- [2] BRYSON, S., KENWRIGHT, D., Cox, M., Ellsworth, D., Haimes, R.: Visually Exploring Gigabyte Data Sets in Real Time, *Communications of the ACM*, vol. 42, No. 8, 1999, 82-90, ACM New York
- [3] EINAV, L., LEVIN, J. D., 2013, The Data Revolution and Economic Analysis, *Working Paper Series*, No. 19035, National Bureau of Economic Research : Cambridge
- [4] LANEY, D.: *Application Delivery Strategies. 3D Management: Controlling Data Volume, Velocity, and Variety*, META Group, File: 949, 6 February 2001. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [date of access: 2 August, 2016]

- [5] GANTZ, J., REINSEL, D.: *The Digital Universe in 2020: Big data, Bigger Digital Shadows, and Biggest Growth in the Far East*, 2012, IDC.
- [6] JESKE, M., GRUNER, M., WEISS, F.: *Big Data in Logistics. A DHL perspective on how to move beyond the hype*, DHL Customer Solutions & Innovations, 2013. Available at: [http://www.dhl.com/content/dam/downloads/g0/about\\_us/innovation/CSI\\_Studie\\_BIG\\_DATA.pdf](http://www.dhl.com/content/dam/downloads/g0/about_us/innovation/CSI_Studie_BIG_DATA.pdf) [date of access: 5 August, 2016]
- [7] Integrated urban e-ticketing for public transport and touristic sites. Science and Technology Options Assessment. European Parliamentary Research Service, Brussels (2014), Available at: [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/513551/IPOL-JOIN\\_ET%282014%29513551\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/513551/IPOL-JOIN_ET%282014%29513551_EN.pdf) [date of access: 5 August, 2016]
- [8] PELLETIER, M.-P., TREPANIER, M., MORENCY, C.: Smart Card Data Use in Public Transit: A Literature Review. *Transportation Research Part C: Emerging Technologies*, vol. 19, No. 4, 2011, 557-568, Elsevier 2011
- [9] URBANEK, A.: Pricing Policy after the Implementation of Electronic Ticketing Technology in Public Urban Transport: An Exploratory Study in Poland, Mikulski, J. (ed.) TST 2015. CCIS, vol. 531, 322-332, Berlin Heidelberg, Springer 2015
- [10] URBANEK A., Electronic Ticket as a Key Element for Integration of Regional and Local Public Transport: Examples of European Solutions, *Logistyka*, No. 4, 2015, 1543-1552, Instytut Logistyki i Magazynowania
- [11] Integrated mobility with eTicketing. The intermodal and interoperable eTicketing platform for consistent, end-to-end mobility chains in cities. Siemens AG (2012), Available at: [http://w3.siemens.com/topics/global/en/events/hannover-messe/metropolitan-solutions/Documents/eTicketing\\_EN.pdf](http://w3.siemens.com/topics/global/en/events/hannover-messe/metropolitan-solutions/Documents/eTicketing_EN.pdf) [date of access: 10 August, 2016]
- [12] SKUP Website [www.kartaskup.pl](http://www.kartaskup.pl) [date of access: 10 August, 2016]
- [13] UNITED NATIONS, Department of Economic and Social Affairs, Population Division, 2014. World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)
- [14] DYDKOWSKI G., URBANEK A.: Directions and Benefits of Using Traffic Modelling Software in the Urban Public Transport, Mikulski, J. (ed.) TST 2013. CCIS, vol. 395, 23-31, Berlin Heidelberg : Springer 2013
- [15] URBANEK A.: Directions of Broadening the Range of Tariff Systems Functionalities in Urban Transport, *Archives of Transport System Telematics*, vol. 7, No. 4, November 2014, 50-54
- [16] DEMPSEY, S. P.: *Privacy Issues with the Use of Smart Cards*. Legal Research Digest, Transport Cooperative Research Program, 2008
- [17] DO, TMT., GATZIA-PEREZ, D.: Where and What: Using Smartphones to Predict Next Locations and Applications in Daily Life, *Pervasive and Mobile Computing*, vol. 12, 2014, 79-91
- [18] DEAKIN, E., KIM, S.: Transportation Technologies: Implications for Planning. University of California Transportation Center, Paper # 536, 2001.