 DOI: 10.22616/rrd.24.2018.072

THE MODEL OF AUTOMATION AND EXTENSION OF TOURISM ECONOMIC IMPACT ASSESSMENT IN SPECIFIC REGIONS

Iluta Berzina, **Ieva Lauberte**
Vidzeme University of Applied Sciences, Latvia
iluta.berzina@va.lv; ieva.lauberte@gmail.com

Abstract
Tourism is measured by the statistics of visitor movements, expenditure and estimates of the number of visitor facilities. In the world there are many current tools, methodologies and innovative technologies used to measure the economic impact of tourism. Tourism statistics have been on the frontline of Big Data-related innovations of statistical sources and methods. Data from mobile phones (as part of Big Data set) are increasingly used as new indices for social science research. Therefore, this paper looks for an answer to the question – what is the specifics of a theoretical model for automation tourism economic impact assessment in specific regions via the use of ICT and mobile positioning data (MPD)? Using qualitative research methods authors propose theoretical model based on two interlinked parts. First of which – 'Data storage' – can be built upon the most popular BD platform 'Apache Hadoop Ecosystem', where the data precision of the online surveys can be increased by implementing mobile positioning solutions. Meanwhile the other part – 'Data analysis' – can be based on the locally created assessment methodology, which has been derived from the Finnish standardized economic impact estimation approach. The research results show the findings and propose a theoretical model. Its strength and novelty lies in the ability to use traditional tourism statistics, the economic impact analysis and passive mobile positioning data for spatial characteristics of tourism flow. Its construction is a distinctive combination of typically used technological approaches.
Key words: tourism economic impact, automation, mobile positioning data.

Introduction

Tourism is a social, cultural and economic phenomenon which entails the movement of people to countries or places outside their usual environment. It is measured by the statistics of visitor movements and expenditure (demand) and estimates of the number of visitor facilities (supply). International tourist arrivals in the world grew by a remarkable 7% in 2017 to reach a total of 1,322 million. This is well above the sustained and consistent trend of 4% or higher growth since 2010 and represents the strongest results in seven years. Based on current trends, economic prospects and the outlook by the United Nations agency World Tourism Organization (UNWTO) and the UNWTO Panel of Experts, UNWTO projects international tourist arrivals worldwide to grow at a rate of 4% – 5% in 2018. This is above the 3.8% average increase projected for the period 2010 – 2020 by UNWTO in its Tourism Towards 2030 long-term forecast (Risi, 2018). Whereas, according to the Central Statistical Bureau of Latvia, the latest statistics show that in 2016 the travellers' balance (exports / imports) remained positive, activity in local recreational trips also increased by 2.1% (CSBL, 2017). Several impacts of tourism can be seen across various sectors, in addition, it is evaluated not only within administrative areas, but also in specific regions such as the National Parks (NPs) (Berzina, 2012). Estimated 311 million people visited NPs in the United States of America (USA) in 2017, visitors spending an added value of approximately 2,15 billion dollars in 2016 (Statista Ltd., 2018). However, according to Eagles (2013), there are significant research gaps that urgently need additional work including the economic impact of park tourism (Eagles, 2013). It is also relevant in Latvia, as the only broader study evaluating the economic significance of tourism in NP regions of Latvia was implemented in 2012, estimating that in 2010 the economic significance of tourism in NP regions of Latvia was more than EUR 71.3 million (Berzina, 2012). In order to ensure EI (economic impact) monitoring at a regional level, in some parts of the world, for example in the USA and Finland, the calculation process is fully or partially automated. There are very many current tools, insights, methodologies and innovative technologies used to measure the impacts of tourism across several sectors and development outcomes (Otarra, 2014). However, the rapid development of information and communication technology (ICT) is changing the research methods or approaches (Raun, Ahas, & Tiru, 2016). Tourism statistics have been on the frontline of Big Data-related innovations of statistical sources and methods (EUROSTAT, 2017). *Big Data* (BD) is high-volume, high-velocity and/or high variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Gartner Inc., 2018). According to the European Union (EU) Agency for Network and Information Security (ENISA), this definition points out the three most outlined dimensions of BD (also known as the 3Vs3) – volume, velocity and variety (ENISA, 2015). Besides these three 'core' Vs of BD, which reflect a more ICT oriented perspective, other key Vs have entered the debate in recent years – veracity, validity, volatility and value (EUROSTAT, 2017). The sources of BD are mobile network operator (MNO) data, smart mobile devices, cameras, Internet of things (IoT) devices, traffic counters and sensors, bank/credit card transitions, web portals and websites, including social media and networks, databases, servers and others. BD when captured, formatted, manipulated, stored and then analysed, can help to gain useful insight to increase revenues, get or retain customers and improve operation, find new correlations of development trends. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s, as of 2012 every day 2.5 exabytes (2.5×1018) of data are generated (Hilbert & Lopez, 2011; IBM, 2012). It is expected that the number of interconnected devices will reach 50 billion by 2020 (Mashal *et al.*, 2015). By 2025, the IDC Corporate (from the USA) predicts there will be 163 zettabytes of data (Reinsel, Gantz, & Rydning, 2017).

Many sources of BD, including mobile positioning data (MPD), are some of the most promising ICT data sources for measuring the mobility of people. Perhaps, the MNO data is the most commonly used BD source for measuring tourism flows. The growing penetration of mobile phone use and falling roaming rates in certain parts of the world (in particular the EU) make the analysis of the whereabouts of mobile phone use a highly relevant source for analysing the presence and movements of tourists (EUROSTAT, 2017). Some researchers consider the results show that MPD has advantages: (1) data can be collected for larger spatial units and in less visited areas, (2) spatial and temporal precision is higher than for regular tourism statistics. Random IDs allow to study tourists' movements (typical routes of certain nationalities) (Ahas *et al*., 2008). Mobile phone data have many unique features and advantages which attract scholars from various fields to apply them to travel behaviour research, and a certain amount of progress has been made to date. However, this is only the beginning, and mobile phone data still have great potential that needs to be exploited for new spatio-temporal tools for improving tourism development planning. Data from mobile phones are increasingly used as an innovative tool in geography and social sciences research (Steenbruggen, Tranos, & Nijkamp, 2015; Wang, He, & Leung, 2017). It gives an opportunity to put forward the *research question* – what is the specifics of a theoretical model for automation tourism economic impact assessment in specific regions via the use of ICT and mobile positioning data (MPD)? In that context authors define the main *research tasks* – (1) to carry out research of the scientific and practical application practice in the world, (2) to evaluate and apply specific technologic combination approach which would be appropriate for the situation in Latvia, as well as to achieve the *aim of the research* – the development of a theoretical model for automating economic impact assessments of tourism in specific regions. At the same time it is also the *novelty of the study*.

Materials and Methods

The study is a qualitative research, therefore it is based on scientific literature and practical findings. The study uses the monographic, comparison, abstract-logical methods; synthesis and analysis as methods are also used. The conclusions of professional organizations, researchers and practical research results relating to the experience of the estimation of tourism economic impact assessment, its automation and extension of European, Asian, American countries (especially the USA, Finland, Estonia) have been studied, selected and used. The methodologies, indicators, data types and sources, technical solutions for automated calculations, potential, problems and limitations have been studied.

There are several economic analysis methods in NP tourism. For instance, based on Burchell and Listokin (1978), Walsh (1986), Warnell (1986), Johnson and Thomas (1992), Williams (1994), Frechtling (1994), *et al.*, theories, which help to estimate extended impact: economic impact (EI), fiscal impact, financial, demand, cost benefit (C/B) analysis, feasibility study, environmental impact and tourism income multiplier assessment (Berzina, 2012). Only a few of them – I/O, C/B, EI analysis and multiplier assessment – are included in the automated or partially automated technical solutions.

In 1976 in the United States Forest Service a linear programming model was created – Impact Analysis for Planning (IMPLAN). IMPLAN estimated the cross-sectoral economic effects of resource outputs on local communities. Now it is developed into the software of an economic impact assessment modelling system, mainly for evaluating development scenarios. This software and exclusively provided several accompanying databases allow building just Leontief's input-output (I/O) models of regional economies. I/O is commonly used to estimate the impact of an economy and to analyse resulting effects that rely on regional economic base data, and an analyst's collected information on a specific economic change of a particular region. Economic consequences can be estimated in the form of jobs, revenues, profits, earnings and/or taxes (IMPLAN, 2018). Model can be used only in administratively defined territories within the region.

The 'Money Generation Model' (MGM) was developed in 1995 and updated in 2000. It is an economic assessment tool available to NP managers in the USA to help gauge the economic impact of NP visitor spending on local economies. The MGM2 estimates direct, indirect, and induced economic effects of visitor spending, and multipliers. Inputs can come from a variety of sources – typically provided by the NPs Public Use Statistics Office, and from NPs Visitor Services Project survey data. If data are not available, generic estimates are provided. EI calculated by the MGM2 are reported in four key areas: sales, jobs, personal income, and value added. The MGM2 model is an Excel based tool (Fish, 2015).

Based on MGM2 model, a method has been developed for standardized EI estimation done by the Finnish Forest Research Institute and Metsähallitus Natural Heritage Services. It is based on the standardized visitor monitoring and provides comparable results between the NPs and other nature recreation areas, and over time. It also enables annual follow-up of the impacts in a cost-effective way. The methodology of tourism impact analysis include the I/O (as IMPLAN) and tourism satellite accounts (TSA) – widely used in the state-level examinations. Finnish standardized EI estimation is done by an Excel-based application 'Paavo'– built on Excel sheets, and it applies Excel functions, macros and SQL queries. The number of visits as well as all the data from visitor surveys is enquired half-automatically from the visitor information database system 'ASTA'. The park classification and multipliers are built in the Excel (Huhtala, Kajala, & Vatanen, 2010). The weakness of the model is its half-automation; the reliability of the method is highly dependent on the success of visitor counting and visitor surveys because the errors in visitor monitoring will be repeated when the total effects are calculated; it can be used just in NPs or other administratively defined territories within the region; access to TSA is needed.

Due to the lack of dynamics of output data, many of EI assessing methods still cannot be used in specific regions of Latvia, except the EI analysis. By combining several techniques used by MGM2 and Finnish standardized economic impact estimation, a methodology for assessing the economic significance of tourism in NP regions was developed, scientifically approved and approbated in Latvia in 2012 (Berzina, 2012). The tourism EI indicators used in the methodology also correspond to the indicators included in the European Tourism Indicator System (ETIS) developed later by the European Commission (EC) in 2016 (EC, 2016). EI analysis estimates the economic significance by calculations of direct, caused, indirect, total economic impacts, and regional economic significance by administrative territories included in the region (Berzina, 2012). The calculation process was not automated; MS Excel and IBM SPSS programs were used for summarizing and analysing the data from surveys. The model has a significant positive aspect – it can be applied even if the research area is not limited to an administrative territory. However, like in Finland, the main drawback is the dependence of the data obtained from surveys on the accuracy of counting. This means additional solutions should be found in order to provide this.

According to de Jonge, Pelt and Roos (2012), there is an additional indicator of economic activity in a region – mobile phone calls, because these data might lead to an indicator that shows that economic activity is changing (Arhipova *et al.*, 2017). The latest research on MPD was carried out in Latvia in 2017. Its methodology corresponds to the field of economics, however, it uses tourism-relevant indicators – mobile call activity, call date, time to test MPD suitability for a theoretical model of updatable Latvian regional business index (Arhipova *et al.*, 2017). One of the conclusions of this study is worth taking into account - the use of MPD allows for a more precise volume determination of the entire assembly of the research and the spatial manifestation of the human flow (directions, intensity).

MPD characterizes the location and movement of a mobile device and it can be divided into two main methods for obtaining the MPD: (1) active positioning and (2) passive positioning. For the active positioning, a specific targeted request can be made to locate the mobile, while for the passive positioning – historical data can be collected and no active requests should be sent. Active and passive data can be monitored both in real-time and historically, although in the active positioning location requests have to be made regularly over a period of time to get the historical data. The active mobile positioning also refers to the ability of installed application to use the location of the mobile device. Mobile applications mostly use three types of location methods – Global Positioning System (GPS)/Assisted Global Positioning System (A-GPS), Wireless networks (WiFi) and network antenna-base location databases. The advantage of the active positioning data is that the geographic information is generally very precise and accurate. The active positioning can be used by researchers for replacing or supplementing travel diaries in spatial behavioural analyses and for generating mobility statistics. The main weakness of this data source is the need to recruit the respondents resulting in a rather small sample size (Ahas, 2014; Tiru, 2014).

Extensive research of active MPD capture and analysis has been carried out by Tasmanian Sensing Tourist Travel (TSTT) project team in Tasmania, and another research was carried out in the port of Palermo, Italy (De Cantis *et al.*, 2016; Hardy *et al.*, 2017). Both tracking applications were implemented in traditional surveys to obtain information about tourists in specific locations. Both studies pointed out that application-based GPS tracking has a huge potential because it gives rich dataset emerged as a result of combining surveys with the precision of the

GPS capabilities of the device. Despite the significant potential of GPS technologies, there are also limits like an investigated group size, battery life of the device and spatial accuracy in a range of situations. To solve the GPS data accuracy, the data were overlaid with different location-based spatial layers (city map, road, NP boundary, etc.) (De Cantis *et al.*, 2016).

Passive MPD are automatically stored in the memory files of mobile operators for call activities or movements of handsets in the network and can also be stored in the memory files of applications in the mobile devices (Ahas *et al*., 2008; Tiru, 2014). The most common source of data in the case of the passive positioning is Call Detail Record (CDR) and Data Detail Record (DDR), which are automatically saved by using the telephone. CDRs can be stored in binary, Extensible Markup Language (XML) or in plain text format (CSV). DDR and CDR data contains – phone ID, the country of registration of the phone, time of event, location coordinates of event. Application based passive MPD is MNO independent and is stored either in the mobile device or in application provider's central databases (Tiru, 2014). Compared to the active positioning data, the spatial accuracy of passive MPD is much lower and the spatial interval is usually irregular and with longer 'time gaps' (Ahas, Raun, & Tiru, 2014). The MNO antennae are distributed unequally throughout the country and also with different network coverage, therefore there is an unequal spatial accuracy – dense regions such as urban areas and roads with heavy traffic have much higher antennae density than rural areas (Kuusik *et al.*, 2011). The main advantage of the passive positioning method is the cost-effectiveness of obtaining huge amounts of data involving all phone users (Tiru, 2014). The passive mobile positioning enables to observe and measure the duration, timing, density, seasonality and dynamics of visits. Moreover, it also allows distinguishing repeat visitors. The repeat visitors could be segmented by their countries of origin, frequency of visitation, seasonality, etc. For the repeat visit determination, 7 days as the preliminary proxy for a single visit can be used. In addition, the local destinations and events most loved by the repeat visitors and their movement trajectories also could be identified (Kuusik *et al*., 2011). The weaknesses of data are related to problems of accessing data, as operators do not wish to share data and because of the privacy and surveillance concerns. Another problem is also that the data is another quantitative dataset with limited features (Ahas *et al*., 2008).

Privacy issues are the most critical aspect of using MPD, because mobile phones become very intimate objects for users (Kuusik *et al.*, 2011). There are several EU laws like General Data Protection Regulation 2016/679 (GDPR), which will be enforced after 25 May 2018 (its successor 95/46/EC), Electronic Privacy Directive 2002/58/EC and other regulations, for example, Directive 223/2009/EC, which determines statistical confidentiality (EP, EC, 2018). Therefore MPD anonymization is very important. The anonymization refers to the process of modifying personal data in such a way that individuals cannot be re-identified and no information about them can be learned. A perfect anonymization is difficult to achieve in practice without compromising the utility of the dataset and with BD this problem increases due to the amount and variety of data. On the one hand, low level of the anonymization (e.g. mere de-identification by just suppressing direct identifiers) is usually not enough to ensure non-identifiability. On the other hand, too strong anonymization may prevent linking data on the same individual (or on similar individuals) that come from different sources and, thus, thwart many of the potential benefits of BD (ENISA, 2015).

According to De Montjove *et al.* (2013) just 4 spatio-temporal points are enough to uniquely identify 95% of the individuals. The analyzed dataset contained 1.5 million phone users with hourly specified individual geographical locations and with a spatial resolution equal to that given by the carrier's antennas. The authors concluded that the spatial aggregation is achieved by increasing the size of the regions in which the user is known to be during his interactions with the service (De Montjoye *et al*., 2013). A significant work of passive MPD analysis and usage was done by Estonians – 'Positium LBS' in cooperation with the University of Tartu. 'Positium LBS' has developed a special software 'Positium Data Mediator' for MNO system. It collects data about call activities (including SMS, MMS, GPRS incoming and outgoing) of selected roaming service users from the billing memory. After that the data collection is processed and anonymized – replacing the direct identifiers with pseudonymous. The pseudonymous data (pseudonymous ID, time of the call activity, cell ID with the geographical coordinates of the antenna, nationality – the country of origin (contract) of the telephone) are obtained from the operators' systems and transferred to the servers of 'Positium Data Mediator'. The visitor flows in the network cells are interpolated geographically for sampling purposes using accommodation statistics, border crossing data and a questionnaire survey. After sampling procedures, the data processing is completed in 'Positium Data Mediator' and the data are ready for analysis (Kuusik *et al.*, 2011).

For determining the usual environment with MPD, Estonian researchers compared 3 different methods. Density-based spatial clustering of applications with a noise method is a rather universal approach to measure people's usual environment and it is most commonly used to determine anchor points from GPS data, but also works well over a longer time period (such as one month) for mobile CDR data. However, this method is rather weak in describing functionality of movements and using in the areas with a mixed land use. Second is an anchor point method whose algorithm is based on the identification of places regularly visited in particular day/night times (Novak *et al*., 2013; Ahas, Raun, & Tiru, 2014). An observation period of at least one month is required to determine the most likely locations for home, work, education, summer home and transportation channels by considering the locations and times of phone usage, visitation frequency, regularity, and variety. In addition to meaningful locations, 'less meaningful' regular visitation areas or secondary anchor points are also determined. The third method is based on borders of administrative units. Researchers compared the distribution of CDR points on a level of a) local community (1 – 5 km); b) municipality (5 – 30 km), c) county (30 – 80 km). This method has less accuracy and has problems with measuring cross-border activities and selecting an appropriate spatial resolution. The positive side of this method is the compatibility with administrative unit based official statistics (Ahas, Raun, & Tiru, 2014). In preparation for a new GDPR (will be enforced in May, 2018), 'Postium LBS' has added one more layer for 'Positium Data Mediator' called 'Sharemind' – secure computing platform developed by 'Cybernetica AS'(Tiru & Bogdanov, 2017). Source data owners encrypt each record on premises with the 'Sharemind' importer and upload to the 'Sharemind' Application Server cluster (Cybernetica AS, 2009). The 'Sharemind' hosts cannot decrypt the data and therefore need to use module called 'Rmind' for the purpose of analysis. At this moment 'Sharemind' is free for developers and privacy researchers.

Results and Discussion

Taking into account that administratively defined territorial units differ from specific regions, as their regionality is more manifested within the boundaries of influences or interests (and thus also within the limits of available official statistics), the authors cannot adapt the IMPLAN technological solution to the tourism EI assessments by developing it up to the process modelling level (Eagles, 2013; IMPLAN, 2018). However, the experience that would be transferable is the possibility of IMPLAN integrated inter-industry economic assessments. Also The MGM2 model is applied within the boundaries of administratively defined territories that constitute the NPs or their regions. As the range of data necessary for calculating multipliers is still not comprehensive in Latvia, but MGM2 requires it, the authors cannot include this approach in the model at present (Berzina, 2012). However, we should take into consideration that data becoming a commodity the situation in the availability of statistics in the world is changing, and the multiplier measurements do not lose their significance. Yet, implementing the goal of the study, the authors can integrate the assessment methodology developed in Latvia derived from the Finnish standardized EI estimation approach in the automated model of the tourism EI assessment (Huhtala, Kajala, & Vatanen, 2010; Berzina, 2012). By the automation of data acquisition, accumulation, summarizing and analysis, on-line surveys can be developed with a direct and immediate connection to the data bank – database system, addressing the dependency of the data obtained from surveys on counting accuracy. This can be increased by the use of MPD (Arhipova *et al.*, 2017). The main strengths of MPD is the spatial accuracy and timeline, which allows a more precise identification of the number of trips taken, number of nights spent, including non-registered accommodation, the duration of the visit, frequency of visits, number of unique visitors, and visited places in the country or region compared to the traditional surveys - based on travellers' honesty and memory. It has also been pointed out that MPD can be used as a supplement for tourism statistics and not as a replacement source of data due to the lack of information about the purpose of a trip and type of accommodation (Ahas, Raun, & Tiru, 2014). Comparing the active positioning with the passive positioning data collection method, there are several differences. The application based active position uses GPS location services, which are independent on a mobile network operator and the data is stored in the mobile device, therefore this method can be used in small sample groups. The main advantages of the passive positioning dataset are costs and speed when obtaining a huge amount of data. The spatial accuracy of passive MPD is lower than in active, but in both cases, as research has showed, there is a need to use overlaying layers not only to correct spatial accuracy, but also to tie with administrative borders. Given that the active positioning dataset is relatively small and, in the specific case, would be equivalent to surveys, MNO provided passive MPD is in use. This has been taken into account in the development of the theoretical model for automating the tourism EI assessment in specific regions of Latvia, and for supplementing these assessments with the spatial dimension of the tourism flow (Figure 1).

The model has been developed based on the assumption that MNO provided sample set of data is at least on one calendar month, it is anonymized and geographically aggregated. Based on the findings of 'Positium LBS', MPD processing is planned to include data cleansing – cleaning from unnecessary

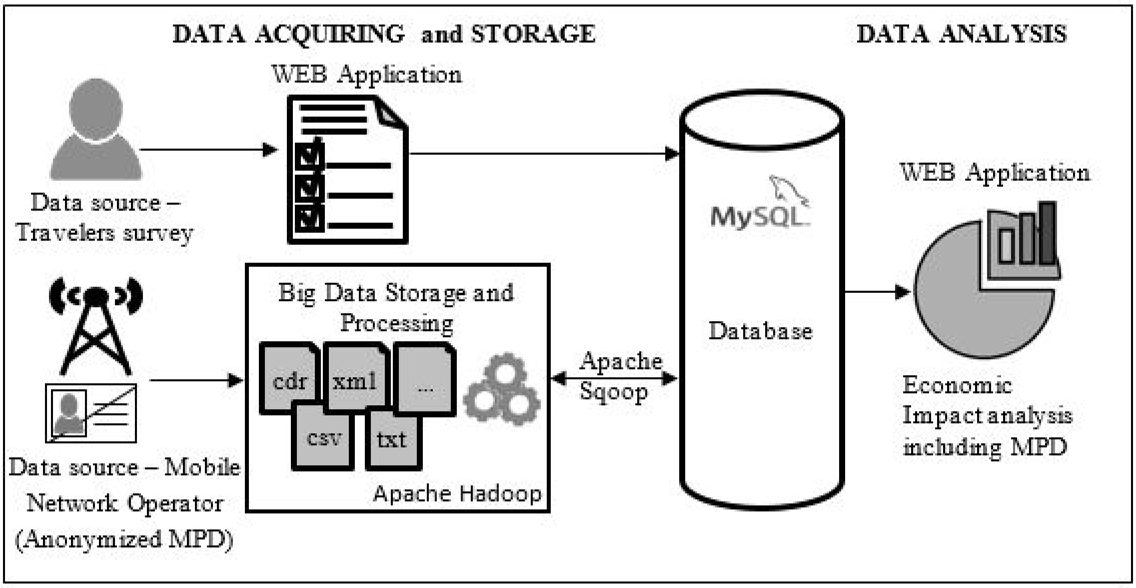

Figure 1. The theoretical model for automating tourism economic impact assessment, and for supplementing it with the spatial dimension (created by the authors).

information, foreign long term visits, cross-border noise, corrupted data, etc., spatial aggregation – spatial interpolated with administrative units – cities and parishes, spatio-temporal statistics – to identify tourist activities and visitor segmentation (Tiru & Bogdanov, 2017). For data storage and processing, the most popular BD platform 'Apache Hadoop Ecosystem' can be used (Apache Software Foundation, 2018). The reason of that decision points to: (1) 'Apache Hadoop' is an open-source software framework that allows for the distributed processing of large datasets across clusters of computers using simple programming models, (2) it is designed to work with any data types – structured, unstructured, semi-structured, which makes it very flexible, (3) in this platform it is possible to use one of the most popular free programming languages and environments called R, which is widely used for statistical analyses and visualizations, (4) the 'Apache Hadoop' software platform is also available as a service from the public cloud providers such as Google Cloud Platform, Amazon Web Services (AWS), Microsoft Azure, Cloudera and others, (5) in the future it will be possible to move to one of the public cloud platforms for more computing resources.

Conclusions

1. The assessment of tourism EI in the USA and Finland is based on the use of fully or partially automated technical solutions, but an essential precondition for their use is that the territory to be assessed must have administrative boundaries.
2. MPD as a part of the BD set has a high potential for the tourism EI assessment. It is also used increasingly in the world to supplement other tourism economic assessments, including the spatial dimension. However, this use is aimed at wider administrative territories than a specific region with peculiar economic impact zones.
3. The methodology for the tourism EI assessment in the NP regions developed in Latvia in 2012 can theoretically be automated. Theoretical model has two interlinked parts. First of them – 'Data storage' - can be built upon the most popular BD platform 'Apache Hadoop Ecosystem', where the data precision of the online surveys can be increased by implementing mobile positioning solutions. Meanwhile the other part – 'Data analysis' – can be based on the locally created assessment methodology. The most important novelty would be that a similar combination of methodological and technological solutions could be used in the regions that are not administratively defined, and such combination has not been developed and tested until now.
4. Using MPD, EU directives and regulations that limit the processing and use of data must be strictly observed. This is a significant challenge for the authors in further scientific-practical research on the particular topic.

Acknowledgements

The article is prepared in frame of the PostDoc research project "Integrated Design of Techo-Social Systems: Next Generation of Tourism Monitoring in Latvia" (SEDAs'RL 10.08.2017. Decision No. 9.-14.1/5018; Project No. 1.1.1.2/VIAA/1/16/110) and supported by the European Regional Development Fund (ERDF).

References

1. Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management. 29 (3), 469–486. DOI: 10.1016/j.tourman.2007.05.014.
2. Ahas, R. (2014). Using Passive Mobile Positioning Data for Generating Statistics: Estonian Experiences. Retrieved February 16, 2018, from: https://www.slideshare.net/Tilastokeskus/mobile-positioning-data-statistics-finland.
3. Ahas, R., Raun, J., & Tiru, M. (2014). Defining usual environment with mobile positioning data. Retrieved February 25, 2018, from: https://ec.europa.eu/eurostat/cros/system/files/Ahas-etal_PaperNTTS2.pdf.
4. Apache Software Foundation (2018) What Is Apache Hadoop? Retrieved March 03, 2018, from: http://hadoop.apache.org/#What+Is+Apache+Hadoop%3F.
5. Arhipova, I., Berzins, G., Brekis, E., Kravcova, J., & Binde, J. (2017). The Methodology of Region Economic Development Evaluation Using Mobile Positioning Data. In 20th International Scientific Conference 'Economic and Social Development' 27–28 April 2017 (pp. 111–120). Prague, Croatia: Varazdin Development and Entrepreneurship Agency.
6. Berzina, I. (2012). Assessment of Tourism Economic Significance in the Regions of National Parks of Latvia. Unpublished doctoral dissertation, Latvia University of Life Sciences and Technologies, Jelgava, Latvia.
7. Cybernetica AS. (2009). Sharemind [computer software]. Tallin, Estonia: Cybernetica AS.
8. CSBL (2017). Tūrisma statistikas aktualitātes 2017. Informatīvais apskats *(Topicality of Tourism Statistics 2017. An Informative Review)*. Retrieved February 5, 2018, from: http://www.csb.gov.lv/sites/default/files/nr_29_turisma_statistikas_aktualitates_17_00_lv.pdf (in Latvian).
9. De Cantis, S., Ferrante, M.,Kahani, A., & Shoval, N. (2016, February). Cruise passengers' behavior at the destination: Investigation using GPS technology. *Tourism Management*. 52. 133–150. DOI: 10.1016/j.tourman.2015.06.018.
10. De Montjoye, Y.A., Hidalgo, C.A., Verleysen, M., & Blondel, V.D. (2013). Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific Reports*. 3 (1376). DOI: 10.1038/srep01376.
11. Eagles, P.F.J. (2013). Research priorities in park tourism. *Journal of Sustainable Tourism*. 22 (4), 528–549. DOI: 10.1080/09669582.2013.785554.
12. EC (2016). European Tourism Indicators System for sustainable destination management. Retrieved January 29, 2018, from: http://www.burrengeopark.ie/wp-content/uploads/2015/03/Sustain-Toolkit.pdf.
13. ENISA (2015). Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics. Retrieved March 02, 2018, from: https://www.enisa.europa.eu/news/enisa-news/privacy-by-design-in-big-data-an-overview-of-privacy-enhancing-technologies-in-the-era-of-big-data-analytics.
14. EP, EC (2018). General Data Protection Regulations [Database EUR-Lex]. Brussels: European Union.
15. EUROSTAT (2017). Tourism statistics: Early adopters of big data? Statistical working paper. 1, 1–30. DOI: 10.2785/762729.
16. Fish, T.E. (2015). Assessing economic impacts of national parks. Retrieved January 30, 2018, from: https://www.nature.nps.gov/parkscience/index.cfm?ArticleID=322&ArticleTypeID=28.
17. Gartner Inc. (2018). *Big Data.* Retrieved February 19, 2018, from; https://www.gartner.com/it-glossary/big-data.
18. Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., & Eccleston, R. (2017). Tracking tourists' travel with smartphone-based GPS technology: a methodological discussion [Electronic version]. *Information Technology & Tourism*. 17(7), 255–274. DOI: 10.1007/s40558-017-0086-3.
19. Hilbert, M., & Lopez, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*. 332 (6025), 60–65. DOI: 10.1126/science.1200970. PMID 21310967.
20. Huhtala, M., Kajala, L., & Vatanen, E. (2010). Local economic impacts of national park visitors' spending: The development process of an estimation method. Metlan työraportteja / Working Papers of the Finnish Forest Research Institute. 149. 20 s. ISBN 978-951-40-2224-1; ISSN 1795-150X.
21. IBM (2012). What is big data? Retrieved January 20, 2018, from: https://www-01.ibm.com/software/in/data/bigdata/.
22. IMPLAN (2018). Software for Economic Analysis. Retrieved February 9, 2018, from: http://implan.com/software/.
23. Kuusik, A., Tiru, M., Ahas, R., & Varblane, U. (2011). Innovation in destination marketing: The use of passive mobile positioning for the segmentation of repeat visitors in Estonia [Electronic version]. *Baltic Journal of Management*. 6 (3), 378–399. DOI: 10.1108/17465261111168000.

24. Mashal, I., Alsaryrah, O., Chung, T.Y., Yang, C.Z., Kuo, W.H., & Agrawal, D.P. (2015). Choices for interaction with things on internet and underlying issues [Survey Paper]. *Ad Hoc Networks.* 28, 68–90. DOI: 10.1016/j.adhoc.2014.12.006.
25. Novak, J., Ahas, R., Aasa, A., & Silm, S. (2013). Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of Estonia. *Journal of Maps*. 9 (1), 10–15. DOI: 10.1080/17445647.2012.762331.
26. Otarra, C. (2014). The Importance of Measuring Tourism Impacts. Retrieved January 21, 2018, from: http://www.solimarinternational.com/resources-page/blog/item/185-the-importance-of-measuring-tourism-impacts.
27. Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data [Electronic version]. *Tourism Management*. 57, 202–2012. DOI: 10.1016/j.tourman.2016.06.006.
28. Reinsel, D., Gantz, J., & Rydning, J. (2017). Data Age 2025: The Evolution of Data to Life-Critical. [IDC White Paper. Framingham, MA, US: International Data Corporation]. Retrieved February 10, 2018, from: https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf.
29. Risi, M. (2018). 2017 International Tourism Results: the highest in seven years. (UNWTO Press Release No. 18003). Retrieved January 22, 2018, from: http://media.unwto.org/press-release/2018-01-15/2017-international-tourism-results-highest-seven-years.
30. Statista Ltd. (2018). National park tourism in the U.S. - Statistics & Facts. Retrieved February 21, 2018, from: https://www.statista.com/topics/2393/national-park-tourism-in-the-us/.
31. Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015, May). Data from mobile phone operators: A tool for smarter cities? Telecommunications Policy. 39 (3–4), 335–346. DOI: 10.1016/j.telpol.2014.04.001.
32. Tiru, M. (2014). Overview of the Sources and Challenges of Mobile Positioning Data for Statistics. Retrieved February 24, 2018, from: https://unstats.un.org/unsd/trade/events/2014/beijing/Margus%20 Tiru%20-%20Mobile%20Positioning%20Data%20Paper.pdf.
33. Tiru, M., & Bogdanov, D. (2017). Mobile Big Data Processing Without Compromising Privacy. Retrieved March 02, 2018, from: https://unstats.un.org/unsd/bigdata/conferences/2017/presentations/day2/session2/case3/5%20-%20POSITIUM_CYBERNETICA%20Margus%20and%20Dan.pdf.
34. Wang, Z., He, S.Y., & Leung, Y. (2017). Applying mobile phone data to travel behaviour research: A literature review [In Press, Corrected Proof]. Travel Behaviour and Society. DOI: 10.1016/j.tbs.2017.02.005.