TACTICAL MODEL FOR CONSTRUCTING A PROTOTYPE OF AUTOMATIZED ASSESSMENT OF TOURISM ECONOMIC IMPACT

Iluta Berzina, **Ieva Lauberte**
Vidzeme University of Applied Sciences, Latvia
iluta.berzina@va.lv

Abstract
The authors explore ways to combine the use of traditional and mobile positioning data (as Big Data) in assessing the economic impact (EI) of tourism in specific regional areas in line with the opportunities of the digital age. In order to provide an automated evaluation of the tourism EI, a study is being carried out resulting in technical development of a prototype for two interconnected digital applications – 'Data Bank' and 'Data Analysis' – as a unified solution for research in the e-environment. The study is based on the research work started in 2017 that resulted in the development of a theoretical strategic model for the construction of the prototype. Continuing the research, the aim of this publication is to develop a theoretical tactical (technological action) model needed for the practical development of the prototype, incl. for the identification of prerequisites to develop the Software Design Description (SDD) of the prototype. To this end, the technological capabilities and their collaborative effects have been investigated in line with the requirements included in the prototype Software Requirements Specification (SRS). Performing the qualitative research, the tactical model has been developed, which designates the practical construction of the prototype. It consists of a combination and joint operation of 4 platforms (servers) – Apache Ambari, Apache Hadoop Ecosystem, Drupal and Jupyter Notebook. The strengths of the model, the risky features as well as the future perspectives for the practical implementation of the goal have been discovered during the research.
Key words: tourism economic impact, prototype, automatized assessment.

Introduction

Regional tourism authorities are increasingly interested in regional statistics of specific features of tourism, for instance the necessity for designing policies, the characteristics of tourism flows, the structure of supply and demand, etc. (UNWTO/INRouTe, 2013). Until recently, there was no standard for tourism research in areas that are affected by this inter-branch of the national economy, especially in the regional level assessments (Berzina, 2012). However, in 2013 the United Nations agency World Tourism Organization (UNWTO) developed the Guidelines for Socio-economic Measurement and Analysis of Regional Tourism. Their focus was on 4 main areas, including 'Tourism as an economic sector' and 'Tourism and sustainable development'. Assessment areas were as follows: supply, demand, employment, business demography, seasonality, economics, related infrastructure, social and cultural impact, etc. UNWTO admits that the 'economic impact' (EI) of tourism is a much broader concept than the 'economic contribution', and it must be estimated by applying models (UNWTO/INRouTe, 2013). This is largely in line with the methodology developed in 2012 in Latvia University of Agriculture (now Latvia University of Life Sciences and Technologies) for assessing the economic significance and the impact of tourism in specific regional areas (Berzina, 2012). The difference is that the Guidelines recommend: regional information should be georeferenced for the purpose of promoting the territorial analysis of tourism activity and other aspects. Such data supports the analysis by allowing the information collected in surveys to be presented spatially and mapped geographically

(UNWTO/INRouTe, 2013). In this context, the experience of Estonia and some Asian countries is a vivid example in the world – traditional tourism statistics and spatial measurements are supplemented by the use of mobile positioning data (MPD) – Big Data (BD) (Berzina & Lauberte, 2018). The United Nations International Telecommunication Union (ITU) has concluded that success in BD analysis and cloud computing will accelerate innovation, significantly transform business, governments and societies worldwide (LETA, 2017).

The authors of the article have also studied the possibilities of combining the use of traditional and MPD in the assessment of the EI of tourism by automatizing the methodology developed in 2012 and supplementing it with the spatial dimension. It has to be admitted that the experience of Estonia has been accepted in the research from the beginning – the acquisition of MPD is possible from the towers of mobile network operators (MDT) (i.e. from MNOs). But in the near future, this research aspect could be reconsidered, as Latvia is in the process of searching for a fundamental solution in the potential use of MPD. Namely, on 01.01.2019 the amendments to the Electronic Communications Law (2004, with amendments made until 03.05.2018) came into force in the Republic of Latvia (LR), which stipulate that the Central Statistical Bureau of Latvia (CSB) has the right to use the data at the disposal of MNOs for the provision of statistics. However, there are no regulations of LR Cabinet of Ministers on how to ensure it (LR Saeima (the Parliament), 2018). There are on-going discussions on the development of the regulations and the result is not expected sooner than

a year or a year and a half (Skreija, 2019). Referring to the opinion of LR Data State Inspectorate (DSI), the creator of statistical information (CSB) is obliged, when creating information, to take into account the technological development tendencies in order to ensure the anonymity of statistical information. Taking into account the possibilities of future technologies, data can be obtained from new sources yet untapped in Latvia (Delfi Bizness, 2019). This means that the research on theoretical aspects of the use of MPD should continue, but only the scientifically tested synthesis of MPD should be used for the purpose of drawing conclusions. The first authors' research results were included in an international scientific publication in 2018, which presented research results that show the findings and propose a theoretical strategic model for building the prototype of two interrelated prototype components – 'Data Bank' and 'Data Analysis' (Berzina & Lauberte, 2018). In order to continue the research, also the prototype of a tool suitable for research in the e-environment is in the process of development. Prototype components are expected to be developed based on the theoretical findings discovered previously and during this study. The authors have developed the prototype Software Requirements Specification (SRS). The SRS includes the general requirements of the prototype and specific requirements, including limitations. For example, the prototype should be based on free software or an open source technologies, and its technical solution can be used in the future as a Platform as a Service (PaaS) on one of the cloud computing platforms. The prototype should provide automatized data processing and analysis insofar as possible, incl. the statistical verification of results. Continuing the work, the authors have successively come to choosing the tactics for further research and practical activities.

The aim of the research: Development of a theoretical tactical (technological action) model required for the practical construction of the prototype, incl. for the identification of prerequisites to develop the Software Design Description (SDD) of the prototype. Tasks: (1) to find out the most appropriate solution or a combination of technologies used for data collection, statistical analysis and their interaction, (2) to develop the tactical model for the construction of the prototype of the automatized assessment of tourism economic impact, (3) to find out theoretical strengths, weaknesses and future perspectives of founded solution or combination of technologies, (4) to create conclusions. The object of the research was to assess the technological possibilities in the construction of the prototype, but the subject of the research to determine the effects of the collaboration of technologies.

Materials and Methods

The study is a qualitative research based on applied and scientific literature and practical findings. The study uses the monographic, descriptive, abstract-logical methods, comparative analysis, analysis and synthesis. A significant range of information sources and technical, analytical descriptions are provided by the official websites of software developers. These frameworks were predominantly created in the United States (US), but the geography of their extensions has a peculiarity – open source technology extensions can be created by anyone anywhere in the world, just as they can be shared in the global public space.

Results and Discussion

In general, the prototype must be capable of operating with both traditional data and MPD. Data acquisition and storage is expected to be implemented through the prototype component 'Data Bank', which is going to be developed so that it includes the possibility to conduct surveys of visitors and entrepreneurs of the research area(s) (primary data acquisition) and integrates the external (secondary) data, including MPD. The prototype component 'Data Analysis' must be able to reveal the socio-economic character of travellers and business, to calculate the economic impact of tourism in research areas in specific regional territories, to perform the statistical analysis of results and to present the results in the spatial dimension.

Technologies investigated for the prototype component 'Data Bank'

The continued research by the authors based in this study on the requirements included in the SRS. To find out the most appropriate solution or a combination of them, there are numerous digital builder tools of online forms for surveys investigated.

According to the methodology of EI assessment, the surveys of visitors are to be prepared in three languages – Latvian, English and Russian, but the surveys of entrepreneurs – in Latvian (Berzina, 2012). However, this is not the only criterion in the platform analysis. It is also necessary to consider the number of possible surveys, the duration of a survey, technical possibilities and their limitations, staticality, etc. The free version of the survey creation platform visidati.lv has a restriction on the number of surveys created, the possible number of respondents, as well as on the duration of free use (three months) (VisiDati SIA, 2018). The limitation of the number of the created surveys was also identified for other platforms examined, such as Wufoo, SurveyGizmo, SurveyLegend (SurveyGizmo LLC, 2018; SurveyMonkey Inc., 2018; SurveyLegend AB, 2018). A limit on the number of respondents was identified for almost all analysed platforms except Surveyplanet, Wufoo, Google Forms (Google LLC,

2018; SurveyMonkey Inc., 2018; Survey Planet LLC, 2018). However, Wufoo also has limits on the number of survey questions and the number of fields to be filled out in addition to the limits on the number of possible surveys (SurveyMonkey Inc., 2018). There are no such limitations for Surveyplanet and Google Forms, but there are other limits (Google LLC, 2018; Survey Planet LLC, 2018) For instance, Surveyplanet supports the creation of surveys in 26 languages, but the Latvian language is not among them (Survey Planet LLC, 2018). For some of the platforms, such as TypeForm, SurveyLegend, the language choice option for creating a survey is available only in the paid version of the platform (SurveyLegend AB, 2018; Typeform, 2018) whereas Google Forms is free only for personal use. On the Google Forms platform a survey can be created in all three languages, but there is a major constraint in the survey interface language – there can only be one language, and this option is linked to the Google account configuration language parameter (Google LLC, 2018). While exploring the capabilities of platforms, the authors discovered that only in one of them, Webropol, it is possible to perform statistical analysis tests, but it can only be done using the paid platform version (Webropol Oy, 2018). Therefore, the platform data export capabilities required for the automatizing of data processing were also evaluated. The authors concluded that in most digital builder tools of online forms for surveys data export is only possible in platform paid versions. It is only available for free on the Google Forms, KwikSurveys, TypeForm, Wofoo platforms (Google LLC, 2018; Problem Free Ltd., 2018; SurveyMonkey Inc., 2018; Typeform, 2018). It could be argued that Google Forms would be the best choice for the platform to use, because it has no limit on the number of questions or the number of forms created, and the download of data is possible. However, it has a major drawback – its free version is only available for personal use and has no possibility for downloading data. All these constraints of various kinds hinder the creation of a convenient and static automatized calculation solution for assessing the economic significance and the impact of tourism. From the perspective of the solution sustainability, some of the platforms that were available free of charge no longer exist, other platforms are developed instead, but some have become a paid or subscription service.

Consequently, the authors explored the possibilities of using one of the open source content management systems for the prototype component 'Data Bank'. Given that not only data collection, storage, but also data sorting and analysis are expected to be carried out in the framework of both components of the prototype, additional requirements were imposed on the content management system. These include: (1)

data security, (2) fast operation, (3) flexibility in form content creation, data sorting and display capabilities, (4) multi-user capability with a mechanism of creator rights, (5) ability to transfer the platform to one of the cloud computing platforms in the future. Therefore, the most popular free open source platforms Wordpress, Joomla and Drupal were explored and analysed (Ivanovs, 2018; POX, 2018). All 3 platforms are available as the service from the public cloud providers such as Google Cloud Platform, Amazon Web Services (AWS) and others. WordPress was originally designed as a blogging platform; therefore, it is more suitable for simple webpages like a blog, and performance is one of its weak points. Drupal is a lightweight platform, which is made for fast performance and is more secure by default installation. From these three platforms, Drupal has more features for managing content types, which will be useful for creating data views, and it has a better user permission management (Mening, 2018; Silverman, 2017). By studying the research results by different authors and the comparisons of these 3 digital tools, it can be concluded that Joomla is something in-between both (Mening, 2018; Morris, 2019; Silverman, 2017). Drupal is more of an enterprise content management system (CMS) solution and it is more popular among enterprise websites (Mening, 2018; Morris, 2019). These aspects convinced the authors that the development of the prototype component 'Data Bank' should be based on the open source CMS Drupal whereas for the acquisition and storage of MPD, the open source BD platform 'Apache Hadoop Ecosystem' has been selected since it has many appropriate libraries for managing and processing BD, including the support for statistical tests. These results of the authors' research were included in the scientific publication in 2018 (Berzina & Lauberte, 2018).

Technologies investigated for the prototype component 'Data Analysis'

According to the EI assessment methodology, the prototype component 'Data Analysis' should be able to perform the division of respondents into 13 segments (Berzina, 2012). It is the basis of the structure for further characteristics and calculations included in the assessments. Among the actions envisaged, the prototype should mainly perform the functions of summation and comparison (difference determination), in specific cases – division, multiplication and expression of results in percentage. From the technical point of view, these are not complicated procedures until the moment when the 'Data Analysis' should perform a statistical analysis of the results obtained. At this point the right choice of the programming language, support mechanisms, possibilities of result visualization, appropriate tools and other technical aspects become crucial. To provide

a functionality of automatized statistical verification of assessment results, the prototype component 'Data Analysis' must be able to carry out 7 kinds of statistical calculations, analyses and tests. To realize this aim, the usage of: (1) developing calculation procedures in the database management system (DMS) MySQL for traditional data, (2) developing calculating functions for traditional data using the Drupal native language PHP, (3) developing calculation procedures using the programming language R and/or (4) the programming language Python has been studied.

The MySQL and PHP – the Drupal native language are not so appropriate for the purposes of statistical analysis without a voluminous programming pre-work. The reason is that these kinds of statistical analyses have never been technically integrated in MySQL and PHP before. Until now, not all, but just some of them are integrated. There are several up-to-date publications confirming that MySQL is not suitable for large scale high-quality data analysis, and other solutions/ technologies are better suited for these aims (Diederich, 2018; Levy, 2016; Periscope Data, 2015). The MySQL optimizer is quite limited because it can execute queries using a single thread, but cannot scale among multiple computer processors cores (Diederich, 2018). The PHP language has some mathematical and statistical extensions that contain functions for statistical computations, but part of them are not documented. PHP is powerful, but not enough for statistical analysis (Ciucu, 2014). The 'Data Analysis' must be able to analyse both MPD and traditional data using a single language, solution. The language R is an environment for statistical computing and graphics that was initially written by Ihaka and Gentleman at the University of Auckland in New Zealand (The R Foundation, 2018). Python is a general purpose language and its greatest strength is a huge amount of libraries like SciPy – a Python-based ecosystem of an open source software for mathematics, science, and engineering, including spatial algorithms. For instance, Voronoi diagrams NumPy – the fundamental package for scientific computing with Python, Matplotlib – Python 2D plotting library, Blaze – query data on different storage systems like the Hadoop ecosystem (Spark, Hive) and other libraries (NumPy developers, 2018; Rodrigues & Overholt, 2015; SciPy developers, 2018; The Matplotlib development team, 2018; The SciPy community, 2018). There is an on-going debate in the research environment about which programming language is the most suitable and fastest for data analysis (Dataquest Blog, 2018; Data-Driven Science, 2018; EliteDataScience, 2017; Ray, 2018). Both are being actively developed in the world. As a result, the major constraints faced by users for a long time have been eliminated. The Python language has diminished

the limitations on data visualization possibilities, while the R language – data wrangling (EliteDataScience, 2017). The Python language is becoming more popular in comparison with the R language, and not only because it is a general purpose language, but it is also increasingly used in Data Science and analytics (Piatetsky, 2018). As a result of the investigation, the authors concluded – the programming languages R and Python show the broadest usage for data statistical analysis in the scientific world.

For the purposes of the prototype component 'Data Analysis', there is a need for a tool that would simplify the technical-practical development in view of its complexity. Increasingly, the digital tool Jupyter Notebook is used in scientific research and publications in Data Science (Randles *et al.*, 2017; Unidata, 2019). The Jupyter Notebook is an open source web application that allows creating and sharing documents that contain a live code, equations, visualizations and a narrative text. The usage includes data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, etc. The Jupiter Notebook has support for over 40 programming languages, including the Python and R language and it has integration with Apache Spark from the Apache Hadoop Ecosystem (Project Jupyter, 2019). It must be noted that the Apache Hadoop Ecosystem has its native notebook tool called Apache Zeppelin. However, when exploring the technical capabilities, strengths and weaknesses of both tools, the authors concluded that Jupyter Notebook is more powerful because of a huge list of supported libraries, including more libraries for visualizing data (Project Jupyter, 2019; Zeppelin, 2019). Zeppelin has a better integration with Apache Hadoop because it is a native tool – it has interpreters for many libraries that are supported in the Apache Hadoop Ecosystem. Both tools can run scheduled tasks, but only Jupyter has options to export a developed notebook as a Python code. It will be very useful for the prototype component 'Data Analysis' to be able to integrate already existing data processing procedures at the end of the development process, thereby automatizing data analysis. There are several ways to connect Jupyter Notebook and get data from the BD platform – run alongside the existing Apache Hadoop platform and access data from Spark, install standalone and use additional tools/extensions like Apache Toree, SparkMagic and Apache Livy, etc. Thereby for creating the prototype component 'Data Analysis', the open source web application tool Jupyter Notebook should be used. It: (1) supports more Python libraries, (2) has better data visualization opportunities, (3) can process both data types – traditional data and MPD, (4) data analyses code can be exported as Python scripts which will be very useful for the automating of the data analyses process.
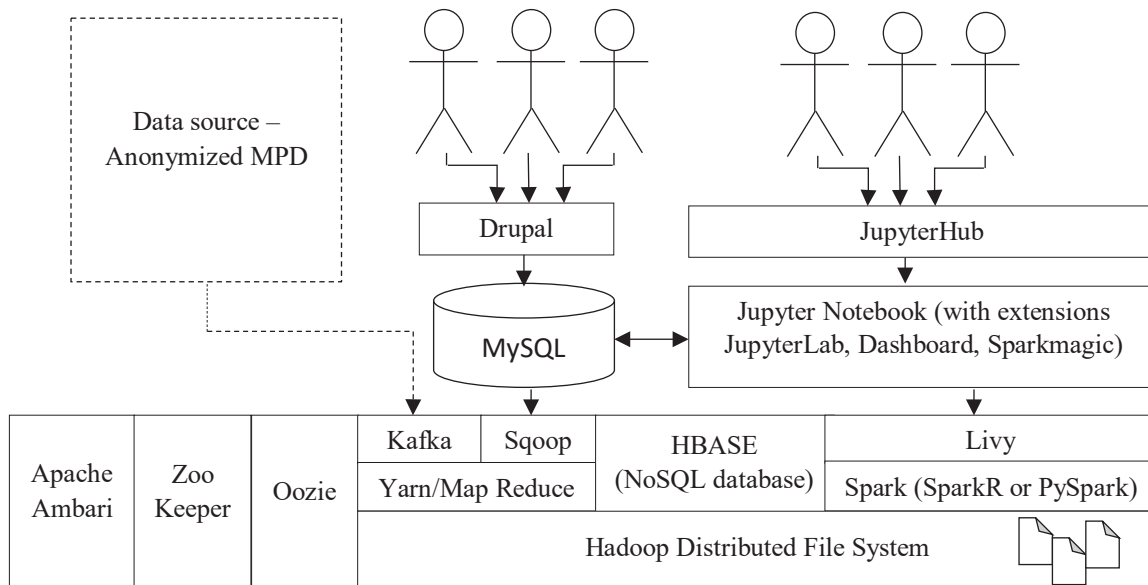
Figure 1. The tactical model for the construction of the prototype of the automatized assessment of tourism economic impact (created by the authors).

Interaction of technologies investigated for the prototype development

Following the examination of the platforms and their support tools, the authors compiled and explained the interconnected technologies and the technologies to be combined that are going to be used in the prototype development. Kafka is a general purpose data collection and streaming tool. For security reasons, Kafka has the encryption of data in-flight using SSL / TLS: This allows data to be encrypted between endpoints (Maarek, 2018). Kafka may be one of the solutions for automatically receiving MPD provided by MNO towers or the LR CSB. Whereas ZooKeeper is like a coordinator that maintains and synchronizes configuration information for distributed applications, Oozie is a workflow scheduler system to manage Apache Hadoop jobs. Yarn and Map Reduce are resource managers for data processing, but HBase is a NoSQL database for MPD storage and data query. Sqoop is like a connector between databases – it can import and export data from MySQL databases. Spark is an analytical framework with programming libraries like SparkR (R on Spark) and PySpark (Python API for Spark). Tool Apache Livy is a connector of Jupyter and Spark (The Apache Software Foundation, 2019). Apache Ambari – a web-based monitoring and managing software for the Apache Hadoop Ecosystems clusters. Jupyther Notebook includes the following extensions: (1) JupyterHub – for multi-user management support, (2) Jupyter Dashboard – layout extensions that allow managing and displaying data analysis results, (3) JupyterLab – enables arranging work area with notebooks and text files in one window, (4) Sparkmagic used for the Jupyter integration with

the BD framework Spark. Drupal is a CMS that in conjunction with the Jupyter Notebook platform and extensions, and the code developed in the Python and/or the R languages can be very powerful for Data Science, because Drupal has the ability to collect information, and Jupyter Notebook – to process and analyse from both data stores – traditional data and MPD.

Based on the investigation results, the authors created the tactical model for the construction of the prototype of the automated assessment of tourism economic impact in specific regions (Figure 1).

Creating a theoretical tactical (technological action) model that is essentially a combination of technological solutions used in the prototype construction, the authors see both its strengths, weaknesses, and future perspectives. The following can be considered as the theoretical values of the model:

1. The ability of the model to be flexible – the wide range of the available library modules in the 'Data Bank' and 'Data Analysis' structure will allow the solution to be used for other research in future;
2. Using the 'Data Bank' content management system, authorized users will be able to easily create new multi-language survey forms, sort the data submitted there, and create new data views;
3. The 'Data analysis' interactive interface will allow easy and fast data analysis and visualization;
4. The standalone system eliminates the risk that a platform manager stops working or makes significant technological changes that threaten the sustainability of the model;

5. After the development and validation of the 'Data Analysis' component of the prototype, it will be possible to export the data analysis stages as executable scripts for further integration into the system as automatic background data processing processes;

6. A model based on open source technologies ensures that there is no need to buy/subscribe to licenses, and the technologies selected provide a possibility for it to be 'transferred' to one of the cloud computing platforms in the future;

7. The technologies support distributed computing.

Risky technological features of the model:

1. The technological combination consists of a set of components, libraries, and it can complicate the technological installation, configuration and administration of the practically implemented prototype.

2. The limitations of the mutual collaboration of libraries are not fully known. Thus, there is a risk of being forced to change the libraries.

Future prospects for the practical implementation and maintenance of the model:

1. To evaluate the integration possibilities of the applications' interfaces, including a unified authentication mechanism, taking into account data security aspects;

2. To assess available resources to separate the externally available information systems from the internal at the network level;

3. To migrate the model to one of the cloud computing platforms.

Conclusions

1. The tactical (technological action) model of the construction of both prototype components –

'Data Bank' and 'Data Analysis' consists of a combination and joint operation of 4 platforms (servers) – Apache Ambari, Apache Hadoop Ecosystem, Drupal and Jupyter Notebook;

2. The model includes promising open source technologies that are combined to produce a powerful framework that can process, analyse and visualize both traditional data and MPD;

3. The tactical model complies with the conditions of the SRS – using free open source solutions that must be able to operate with traditional and MPD, perform simple and more complex calculations, and the statistical analysis and visualization of results. Taking into account that the methodology for evaluating tourism EI is a set of mathematically static (fixed) actions, the requirement for ensuring a higher statisticality of the model is also met – considering the peculiarities of cooperation of all involved technologies, it is possible to transfer the model to one of the cloud computing platforms;

4. In the course of the research, the values of the model, possible risky features and future perspectives for the implementation of the model have been identified. This proves that there is a basis for the continuation of the research, but already in parallel with the technical development of the prototype.

References

1. Berzina, I. (2012). *Assessment of Tourism Economic Significance in the Regions of National Parks of Latvia.* Unpublished doctoral dissertation *(Full text)*, Latvia University of Agriculture, Jelgava, Latvia.

2. Berzina, I., & Lauberte, I. (2018). The Model of Automation and Extension of Tourism Economic Impact Assessment in Specific Regions. In Annual 24th ISC Research for Rural Development 2018 Vol. 2, 16-18 May 2018 (pp. 195–202). Jelgava, Latvia: Latvia University of Life Sciences and Technologies. DOI: 0.22616/rrd.24.2018.072.

3. Ciucu, S.C. (2014). Statistical Data Analysis via R and PHP: A Case Study of the Relationship Between GDP and Foreign Direct Investments for The Republic Of Moldova. Romanian Statistical Review No. 2, 2014. Retrieved December 15, 2018, from http://www.revistadestatistica.ro/wp-content/uploads/2014/07/RRS_2_2014_a03.pdf.

4. Data-Driven Science. (2018, January 31). Python vs R for Data Science: And the winner is. Retrieved April 25, 2018, from https://medium.com/@data_driven/python-vs-r-for-data-science-and-the-winner-is-3ebb1a968197.

5. Dataquest. (2018, November 1). Python vs R: Head to Head Data Analysis. Retrieved January 14, 2019, from https://www.dataquest.io/blog/python-vs-r/.

6. Delfi Bizness (2019, February 25). Datu valsts inspekcija: statistikā var izmantot neidentificējamu personu datus no mobilajiem operatoriem (Data State Inspectorate: Data of unidentifiable individuals from mobile

operators can be used in statistics). *Delfi Bizness*. Retrieved February 25, 2019, from https://www.delfi.lv/bizness/biznesa_vide/datu-valsts-inspekcija-statistika-var-izmantot-neidentificejamu-personu-datus-no-mobilajiem-operatoriem.d?id=50859349. (in Latvian)

7. Diederich, T. (2018, January 5). 5 Limitations of MySQL with Big Data, Friday. *GridGain*. Retrieved April 25, 2018, from https://www.gridgain.com/resources/blog/5-limitations-mysql-big-data.

8. Elite Data Science (2017, February 11). R vs. Python for Data Science: Summary of Modern Advances. Retrieved April 25, 2019, from https://elitedatascience.com/r-vs-python-for-data-science

9. Google LLC. (n.d.). Collect and organize information big & small with Google Forms. For free. Retrieved April 23, 2018, from https://www.google.com/forms/about.

10. Ivanovs, A. (2018, May 30). The 9 Most Popular Free Content Management Systems (CMS) 2018. *Colorlib*. Retrieved December 15, 2018, from https://colorlib.com/wp/most-popular-content-management-systems.

11. LETA (2017, November 21). Latvija 35.vietā pasaulē informācijas un komunikāciju tehnoloģiju attīstībā (Latvia takes the 35th place in the world in the development of information and communication Technologies). *Dienas Bizness*. Retrieved February 22, 2019, from https://www.db.lv/zinas/latvija-35-vieta-pasaule-informacijas-un-komunikaciju-tehnologiju-attistiba-469046. (in Latvian)

12. Levy, E. (2016, December 15). Pros and Cons of Using MySQL for Analytical Reporting. *Sisense*. Retrieved December 15, 2018, from https://www.sisense.com/blog/pros-cons-using-mysql-analytical-reporting/

13. LR Saeima (the Parliament) (2018, May 3). Elektronisko sakaru likums (Electronic Communications Law). *Latvijas Vēstnesis*. Retrieved February 25, 2019, from https://likumi.lv/doc.php?id=96611. (in Latvian)

14. Maarek, S. (2018, March 13). Introduction to Apache Kafka Security. *Medium*. Retrieved January 10, 2019, from https://medium.com/@stephane.maarek/introduction-to-apache-kafka-security-c8951d410adf.

15. Mening, R. (2018, June 1). WordPress vs Joomla vs Drupal (Comparison). *WebsiteSetup*. Retrieved April 27, 2018, from https://websitesetup.org/cms-comparison-wordpress-vs-joomla-drupal/.

16. Morris, W. (2019, January 19). Joomla vs Drupal: Two Top Content Management Systems Compared. *Hostinger Tutorials*. Retrieved February 5, 2019, from https://www.hostinger.com/tutorials/drupal-vs-joomla.

17. NumPy developers. (n.d.). NumPy. Retrieved December 17, 2018, from http://www.numpy.org/.

18. Periscope Data. (2015, June 18). 4 Reasons Not To Use MySQL For Analysis. Retrieved April 25, 2018, from https://www.periscopedata.com/blog/reasons-not-to-use-mysql.

19. Piatetsky, G. (2018, May). Python eats away at R: Top Software for Analytics, Data Science, Machine Learning in 2018: Trends and Analysis. *Knuggets*. Retrieved January 15, 2019, from https://www.kdnuggets.com/2018/05/poll-tools-analytics-data-science-machine-learning-results.html.

20. POX. (2017, November). TOP 10 open source CMS for building a website. *AlternativeTo*. Retrieved April 27, 2018, from https://alternativeto.net/list/341/top-10-cms-for-building-a-website.

21. Problem Free Ltd. (n.d.). A survey maker, for individuals and teams. Retrieved April 23, 2018, from https://kwiksurveys.com/.

22. Project Jupyter. (n.d.). Documentation. Retrieved January 10, 2019, from https://jupyter.org/documentation.

23. Randles, B.M., Golshan, M.S., Pasquetto, I.V., & Borgman, C.L. (2017). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. 2017 ACM/IEE Joint Conference on Digital Libraries (JCDL) 19-23 June 2017. Toronto, ON, Canada. DOI: 10.1109/JCDL.2017.7991618.

24. Ray, B. (2018, June 10). Python vs (and) R for Data science. *Noteworthly – The Journal Blog*. Retrieved January 14, 2019, from https://blog.usejournal.com/python-vs-and-r-for-data-science-833b48ccc91d.

25. Rodrigues, D., & Overholt, K. (2015, September 16). Analyzing 1.7 Billion Reddit Comments with Blaze and Impala. Retrieved January 20, 2019, from http://blaze.pydata.org/blog/2015/09/16/reddit-impala/.

26. SciPy developers (n.d.). Scientific Computing Tools for Python. Retrieved December 15, 2018, from https://www.scipy.org/about.html.

27. Silverman, M. (2017, January 4). 5 Reasons to use Drupal vs. WordPress. *Duo*. Retrieved April 27, 2018, from https://thoughts.duoconsulting.com/blog/5-reasons-to-use-drupal-vs.-wordpress.

28. Skreija, D. (2019, February, 22). Plāno izmantot mobilo operatoru datus iedzīvotāju skaita un kustības noteikšanai, operatori bažīgi (Mobile operator data is expected to be used to determine the population and its movement, the operators are worried). *Delfi Bizness*. Retrieved February 25, 2019, from https://www.delfi.lv/bizness/tehnologijas/plano-izmantot-mobilo-operatoru-datus-iedzivotaju-skaita-un-kustibas-noteiksanai-operatori-bazigi.d?id=50834631&fbclid=IwAR3UFKtb5ZJtPbkJRaDaH68A-W-5p4cLzq89GZUZZZ0qVZp1-3n_Uu4-YSpg. (in Latvian)

29. Survey Gizmo LLC. (n.d.). Enterprise Survey Features & Licenses. Retrieved April 23, 2018, from https://www.surveygizmo.com/survey-software-features.

30. Survey Legend AB. (n.d.). Features By The World's Leading Brands. Retrieved April 23, 2018, from https://www.surveylegend.com/features.

31. Survey Planet LLC. (n.d.). Free Features. Retrieved April 23, 2018, from https://surveyplanet.com/features#free-features.

32. SurveyMonkey Inc. (n.d.). Ready to get started for FREE? Retrieved April 23, 2018, from https://www.wufoo.com/pricing/.

33. The Apache Software Foundation. (n.d.). Apache Oozie Workflow Scheduler for Hadoop. Retrieved January 3, 2019, from http://oozie.apache.org/.

34. The Matplotlib development team. (n.d.). Matplotlib. Retrieved December 18, 2018, from https://matplotlib.org/.

35. The R Foundation. (n.d.). What is R? Retrieved December 15, 2018, from https://www.r-project.org/about.html.

36. The SciPy community. (n.d.). Spatial algorithms and data structures (scipy.spatial). Retrieved December 20, 2018, from https://docs.scipy.org/doc/scipy/reference/spatial.html.

37. Typeform, S.L. (n.d.). Choose the plan that works for you. Retrieved April 23, 2018, from https://www.typeform.com/pricing/.

38. Unidata Online Python Training. (n.d.). Why Python and Jupyter Notebooks? Retrieved January 20, 2019, from https://unidata.github.io/online-python-training/introduction.html.

39. UNWTO/INRouTe (International Network on Regional Economics, Mobility and Tourism and World Tourism Organization) (2013). *A Closer Look at Tourism: Sub-national Measurement and Analysis – Towards a Set of UNWTO Guidelines*. DOI: 10.18111/9789284414963.

40. VisiDati SIA. (n.d.). VisiDati.lv. Retrieved April 23, 2018, from https://www.visidati.lv/prices/.

41. Webropol Oy. (n.d.). Survey Analytics Software Made Easy. Retrieved April 24, 2018, from http://w3.webropol.com/additional-modules/online-survey-analytics/.

42. Zeppelin. (n.d.). What is Apache Zeppelin? Retrieved February 10, 2019, from http://zeppelin.apache.org/docs/0.8.1/.