SJPPPA
SLOVAK
JOURNAL
OF PUBLIC
POLICY
AND PUBLIC
ADMINISTRATION

Markéta VOGELOVÁ[1] - Radek CHALOUPKA[2] - Josef ABRAHÁM[3]

POTENTIAL OF NEW DATA SOURCES IN TOURISM

Abstract

The intensive development of new technologies extends the possibilities for data resources to monitor tourism. New sources offer a wide range of data that could be used to complement traditional statistical resources. In this context, however, a number of methodological and legislative issues need to be addressed. This paper presents a contribution to the discussion on the use of data sources to measure the performance of road traffic more heavily. The aim of this article is to analyze the potential of modern data sources in terms of their usability for statistics and performance measurement of tourism. In order to determine the potential of new statistical sources, a questionnaire survey of respondents from the experts in the field of tourism is used. The research is mainly focused on the needs of destination management organizations in the Czech Republic. The theoretical framework of the article is an analysis of the available sources of tourism statistics in the Czech Republic and in the international environment.

KEY WORDS: tourism, statistical data, geolocation data, payment card data

1   INTRODUCTION

Traditional tourism statistics have long been based on original data sources, primarily administrative data from national and international organizations. These data are continuously supplemented by thematic research and analysis. The available data allowed a comparison of selected tourism indicators at regional, national or international level. The basic advantage of administrative data is the long-term unified methodological approach in collecting and evaluating and solving legislative issues related to the protection of individual data. Their disadvantage is the limited possibility of obtaining the necessary data on all

1 Ing. Markéta Vogelová, University of Economics, Prague, nám. W. Churchilla 1938/4, 130 67 Prague, Czech Republic, e-mail: vogm00@vse.cz.

2 Ing. Radek Chaloupka, University of Economics, Prague, nám. W. Churchilla 1938/4, 130 67 Prague, Czech Republic, e-mail: char01@vse.cz.

3 Doc. Ing. Josef Abrhám, Ph.D. Metropolitan University Prague, Dubečská 900/10, 100 31 Prague, Czech Republic, e-mail: josef.abrham@mup.cz.

relevant indicators of the tourism industry at an affordable price. Traditional statistics have always provided, to a limited extent, information on the size and performance of the tourism industry. Intense technological development based on the use of modern digital technologies has brought the possibility of using large volumes of data called Big Data. The technological boom was also a rocket growth in Internet platforms and the emergence of new tourism service providers. The development of social networks and new opportunities for marketing / communication activities also significantly affected the tourism sector. The use of modern data sources based on Big Data represents considerable potential in terms of tourism. Statistics in connection with the emergence of new opportunities are in many ways once again at the beginning. It is necessary to address not only methodological problems but also a number of legislative issues. Due to the large number of modern (new) data sources, it is necessary to look at their usability in terms of tourism statistics.

The aim of this article is to evaluate the potential of modern data sources in terms of their usability for tourism statistics and measuring the performance of tourism. The research is mainly focused on the needs of destination management organizations in the Czech Republic. The methodological approach of the solution is based mainly on qualitative methods of market research. In order to determine the potential of new statistical sources, a questionnaire survey of respondents from the experts in the field of tourism is used. The assessment of the potential of modern data sources of tourism statistics is based on research results. The theoretical framework of the article is an analysis of the available sources of tourism statistics in the Czech Republic and in the international environment. Tourism statistics are divided into traditional new data sources. Traditional sources are the statistics of public institutions aagentur in the Czech Republic (CSO, MRD, CNB, CzechTourism) and international organizations (Eurostat, UNWTO, OECD, World Travel & Tourism Council). Modern sources of tourism statistics include geolocation / positioning data of mobile operators, residual data of international payment associations, data from social networks, data of shared economy platforms, and data of internet search engines. The evaluation of the data sources and the identification of the key conclusions were based on the analysis of strengths and weaknesses.

from these entities. However, the disadvantage of execution is that it is relatively time consuming and costly.

2   OVERVIEW OF TRADITIONAL AND MODERN DATA SOURCES IN TOURISM

Traditional data sources are for analysis purposes defined as data, that are commonly available and used for decision-making by tourism stakeholders.

Markéta VOGELOVÁ - Radek CHALOUPKA - Josef ABRAHÁM

MARKÉTA VOGELOVÁ - RADEK CHALOUPKA - JOSEF ABRAHÁM

They include the statistics of the Czech Statistical Office (CZSO), the Ministry for Regional Development (the Ministry), the Czech National Bank (CNB) and the Czech Tourism Authority - CzechTourism.

The Czech Statistical Office is the main source of statistical data in the field of tourism. It deals with the acquisition, processing, analysis and interpretation of statistical data from tourism and travel. In particular, it provides data on the volume of tourism, its characteristics, the profile of tourism participants, tourism expenditure and tourism benefits for the national economy. Data provision is governed by EU legislation and international recommendations on tourism statistics, in particular Regulation EU 692/2011 on European statistics on tourism. The CZSO identifies and provides data primarily in the following key areas: statistical survey at collective accommodation establishments (capacity indicators and occupancy + database management of these facilities); tourism survey (information on domestic and outbound trips of Czech households); creating a tourism satellite account and a tourism employment module.

The Ministry for Regional Development of the Czech Republic, in cooperation with the Czech Statistical Office and the CzechTourism Agency, has been implementing a system of national projects to improve statistical data on tourism. The result was the implementation of statistical censuses in the field of tourism - accommodation, catering and MICE. The Ministry has in the past been the initiator and long-term holder of the project "Inbound tourism".

The Czech Tourism Authority - CzechTourism deals mainly with marketing research including motivational surveys in the field of tourism such as tracking of domestic and inbound tourism, country reports, processing of infographics and analyses. CzechTourism is the implementor of research projects in the field of tourism statistics - Inbound tourism. The Czech National Bank is the provider of data on the balance of payments of tourism.

The source of tourism statistics at international level are data from Eurostat, the World Tourism Organization, the Organization for Economic Cooperation and Development. The importance of these organizations lies in the methodological unification of tourism statistics, international benchmarks and sectoral / intersectoral analysis. Eurostat monitors data on the territory of the European Union, including traffic statistics and tourism capacities. The role of Eurostat is also to unify the level of tourism statistics and methodological support of individual member countries (Eurostat, 2016).

The World Tourism Organization (UNWTO), which is a specialized agency of the United Nations, is considered the most comprehensive source of internationally comparable data in recent years. Data are published in almost 160 member countries of UNWTO. Furthermore, they are monitored not only by countries, but also by larger units (developed and emerging economies, and especially UNWTO regions and subregions). UNWTO monitors tourism statistics

on a global level. It provides a prediction of tourism development and analyzes the current and historical development of tourism. Its role also lies in providing methodological support to member countries (World Tourism Organization, 2018).

Organization for Economic Co-operation and Development (OECD) focuses on analyzing selected issues of tourism in the context of the development of the international economy. It processes thematic analysis of selected countries or international comparisons (Dupeyras, 2013).

Information on the impact of tourism is based on an unofficial world economy satellite account, which is compiled by the World Travel & Tourism Council (WTTC) every year. Unlike the UNWTO methodology, economic impacts are monitored instead of international flows (revenues, expenditures, etc.). In addition, WTTC statistics include domestic tourism data. The ranking of countries in international comparison according to UNWTO and WTTC may vary significantly. Countries with the most arrivals by UNWTO (eg France or Spain) are not dominant in terms of value added or employment in tourism. In terms of the impact on the economy, the US, Japan, China or India have high levels of contrast (World Travel&Tourism Council, 2018).

Only some of the new data sources available for tourism statistics were selected for analysis purposes. The selection was made on the basis of practical experience with the less usable resources, that are used recently for tourism statistics and the measurement of their selected aspects of the industry. From the point of view of the analysis, these resources are divided into five basic groups: geolocation-positioning data (mobile operators), data from payment cards (credit card associations), data from social media (Facebook, Twitter, Youtube, Instagram, (Airbnb, Uber, Liftago, HomeAway, etc.) and Internet search engines (Google, List, etc.).

Geolocation data from mobile operators provide information about the amount and profile of visitors of the destination / area over time. This source also provides information on geographic indicators, such as the number of foreign tourists connecting to the local network, their period of stay in the country, frequency analysis of visiting destinations, etc. The advantage is the possibility of monitoring the trajectory (the whole route) of the foreign visitor's stay. The positioning data provide data for the calculation of aggregated indicators, such as the number of foreign tourists in total, the number of one-day visitors, the number of transits. Due to the personal data protection, this source does not provide an individual identification, including basic demographic indicators (Seynaeve, 2016).

Payment card data enable geographical identification of expenses (information on the place of use of the payment card, the method of its use, payments and withdrawals of cash on a given payment card, etc.) Provide data

Markéta VOGELOVÁ - Radek CHALOUPKA - Josef ABRAHÁM

MARKÉTA VOGELOVÁ - RADEK CHALOUPKA - JOSEF ABRAHÁM

for calculation of average and total expenses, information on places with the highest number of cash withdrawals, places with the highest expenses, etc. They do not allow the identification of the individual, including the basic demographic indicators.

Shared economy data allow to measure the performance of these platforms. In terms of tourism statistics, they are particularly important for the accommodation, catering and transport sectors. These are primarily Airbnb, HomeAway, Vacation Rentals, Homeliday, FlipKey, Uber, Liftago, Taxify, BlaBla cars, etc. Data are not provided by platforms to the extent necessary for use in tourism. Another disadvantage is the fact that there is no relevant methodology for their use.

2.1 Research of the potential of data sources in tourism

In order to determine the potential of new statistical sources, a questionnaire survey of respondents from the experts in the field of tourism is used. The research is mainly focused on the needs of destination management organizations in the Czech Republic. The thematic research focuses on geolocation data, data from payment cards, social networks, shared economy platforms, internet search engines and their comparison with traditional sources. The questionnaire survey was conducted between February and March 2018. 79 tourism experts were questioned. Respondents were geographically distributed within the territory of the Czech Republic. The questionnaire survey was divided into two basic parts. The first part of the research focused on comparing the potential of traditional and new data sources for tourism statistics. The second part of the research was devoted exclusively to new data sources of tourism statistics and the way they are used by the tourism stakeholders and the role of the public sector.

The results of the questionnaire survey are divided into 7 specific areas as follows: 1) evaluation of the importance of selected data sources in terms of the needs of respondents to research in the field of tourism statistics, 2) respondents ' confidence in relation to individual data sources as sources for tourism statistics, 3) reasons to prevent respondents from using individual data sources for statistical purposes in the tourism sector, 4) evaluation of future possibilities and potential use of data sources for tourism statistics, 5) preferences of respondents way of providing statistical data from new data sources for the needs of tourism statistics, 6) respondents ' opinion on the payment of statistical data from new data sources a 7) preferences of respondents to regional breakdown statistical data from new data sources and their provision for public use.

Graph 1 Evaluation of the significance (importance) of selected data sources in terms of the needs of respondents to research in the field of tourism statistics
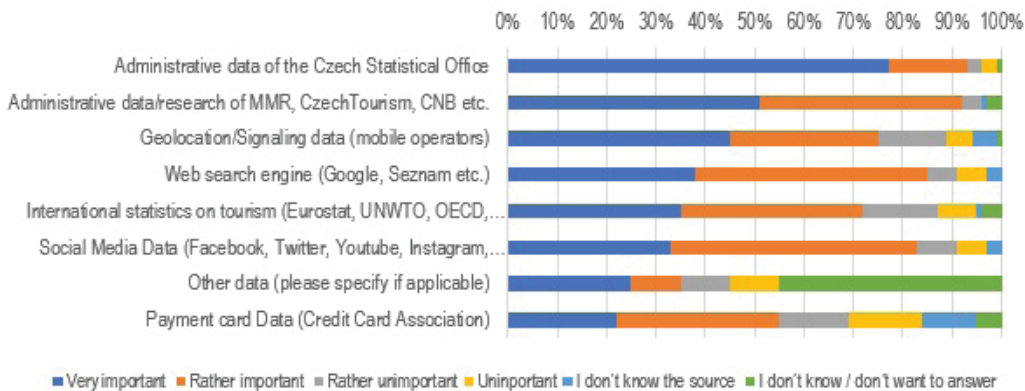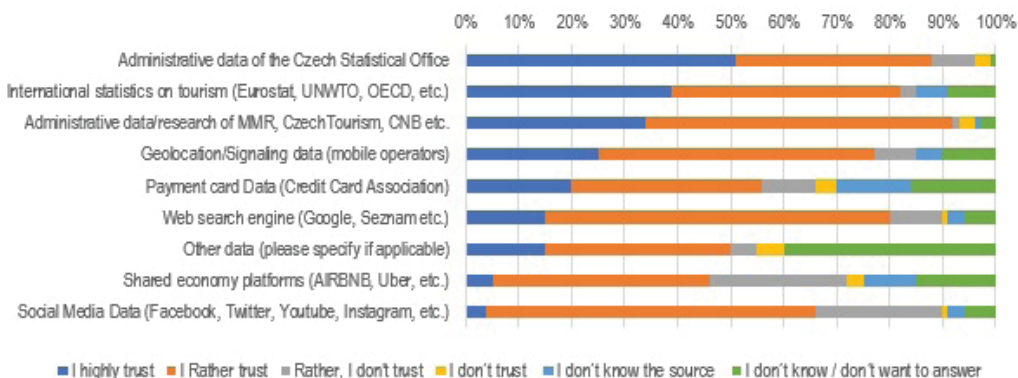

Source: Own research

When it comes to assessing the importance of data resources in terms of tourism needs, respondents prefer the most the administrative data of the Czech Statistical Office, data and research by the Ministry for Regional Development of the Czech Republic, geolocation data of mobile operators and Internet search engines. Data from payment cards and social media are the least significant ones (see Graph 1 for details).

An important factor in using modern data sources is their credibility. As can be seen from Graph 2, the respondents logically more trust the traditional national resources and international statistics. In the case of modern data sources, they perceive mobile positioning data and Internet search engines as more reliable. On the contrary, they trust the credit card and social networking data the least.

Graph 2 Respondents ' confidence in relation to individual data sources as sources for tourism statistics

Source: Own research

MARKÉTA VOGELOVÁ - RADEK CHALOUPKA - JOSEF ABRAHÁM

Concerning the weakness of individual data sources, the questionnaire survey also confirmed the weakness of traditional tourism statistics, such as the unpredictability of the data provided, the limited number of statistical indicators and their insufficient regional breakdown. In relation to the new data sources, experts see the complexity of the data source methodology and the lack of clarity and the level of interpretation of their results (see Graph 2) as deficiencies.

Graph 3 Reasons to prevent respondents from using individual data sources for statistical purposes in the tourism sector
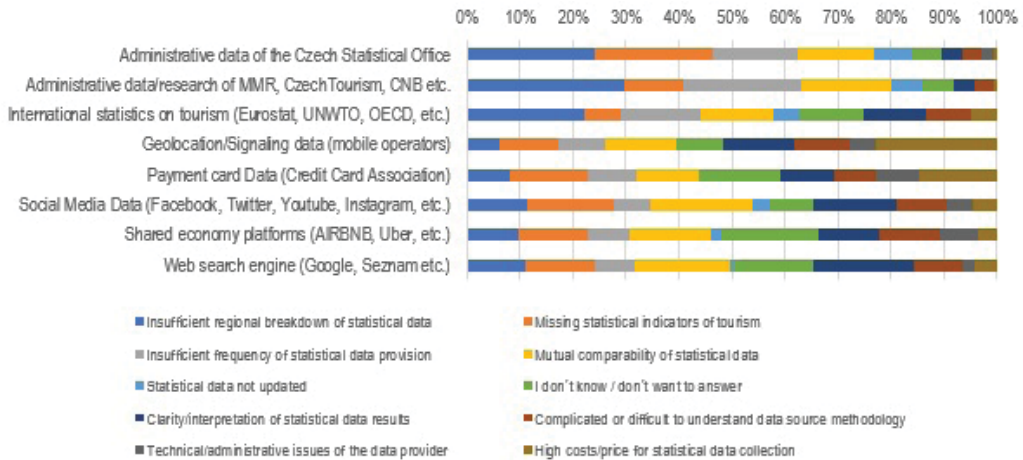
Source: Own research

Graph 4 presents the results of the potential assessment for the future use of individual data sources of tourism statistics. Experts generally perceive all tested sources of tourism statistics as usable. Geolocation / positioning data (89 %) and Internet search engines (84 %) have the biggest potential. High potential also keeps the traditional source of tourism statistics, which are the administrative data of the Czech Statistical Office. On the contrary, experts believe the possibilities of using shared economy data (66 %) the least, which may be related to their relatively low confidence.
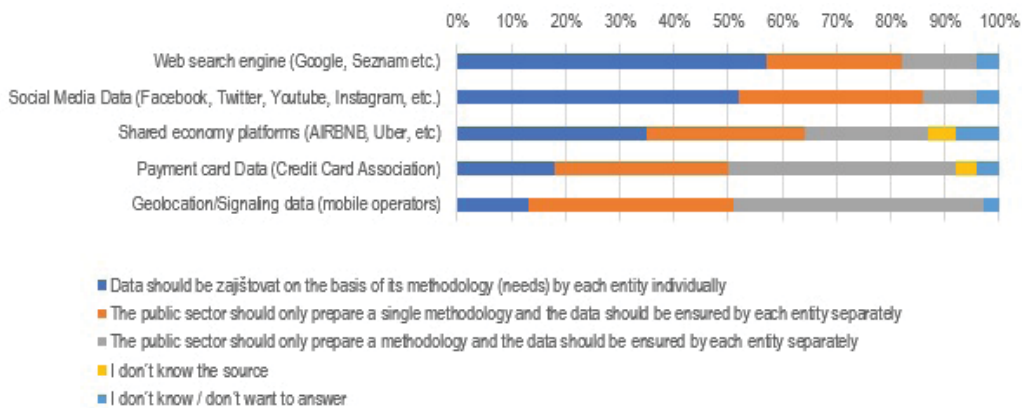
Graph 4 Evaluation of future possibilities and potential use of data sources for tourism statistics
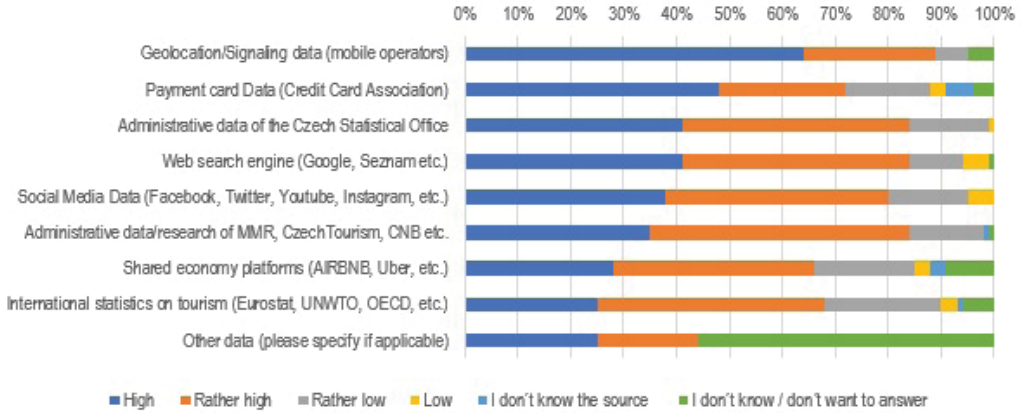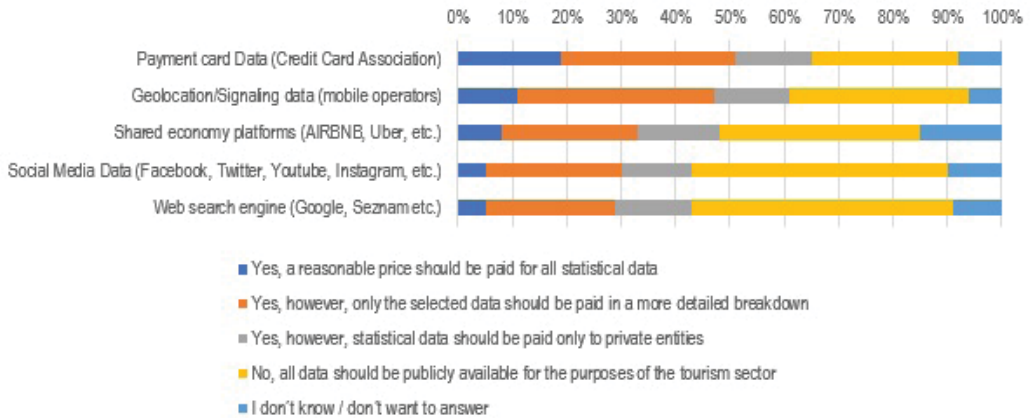

Source: Own research

It is clear from Graph 4 that respondents support the creation of a uniform methodology for the use of new data sources by the public sector. From the point of view of ensuring full data utilization, including uniform methodological unification, the most preferred geolocation / positioning data and payment card data are the most preferred. Concerning data from social networks, the public sector should engage rather in methodological support.

Graph 5 Preferences of respondents way of providing statistical data from new data sources for the needs of tourism statistics
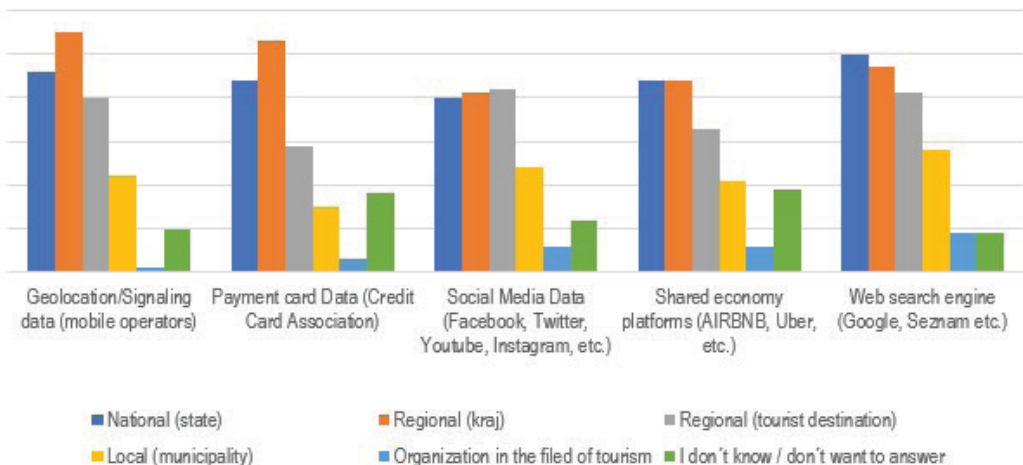

Source: Own research

Markéta VOGELOVÁ - Radek CHALOUPKA - Josef ABRAHÁM

Markéta VOGELOVÁ - Radek CHALOUPKA - Josef ABRAHÁM

Graph 6 Respondents ' opinion on the payment of statistical data from new data sources

Source: Own research

In the survey, respondents were asked whether they would be personally willing to pay for statistical data. A total of 61 % of respondents are willing to pay a reasonable price for the data, while 39 % of respondents are reluctant to pay no costs for the data from new sources. This result is probably due to the fact, that tourism stakeholders already have to pay for a range of data, especially from the field of marketing research. The willingness to pay some financial resources decreases with the origin of the respondent. Significantly decreasing when it is a public sector entity (see Graph 6 for details).

Graph 7 Preferences of respondents to regional breakdown statistical data from new data sources and their provision for public use

Source: Own research

Graph 7 illustrates, that data from new sources should be available publicly and free of charge at the national and regional level, according to the results of the questionnaire survey, with each new data source being different. The largest number of respondents supports the provision of geolocation (signaling) data from mobile operators, payment card data and Internet search engines. With a lower level of tourism management, experts' preferences to data provision are declining.

Following the overall results of the questionnaire survey, priority was given to support the use of new data sources of tourism in this sector. The design of priorities and involvement of the public sector in the support of individual data sources is presented in Table 1. The public sector involvement in the areas of preparation of a unified methodology, regional breakdown and mediation of data availability was considered as supported.

Table 1 Prioritization public sector activities in suporting the exploitation of new tourism data sources

New data source	Preparation of uniform methodology	Level of data availability	Level providing free data
Geolocation/Signaling data (mobile operators)	Very high priority	National, Regional (kraj), Regional (tourist destination), Local	National, Regional (kraj)
Payment Card Data	High priority	National, Regional (kraj), Regional (tourist destination), Local	National, Regional (kraj)
Shared economy platforms	High priority	Micro level	National
Social Media Data	High priority	Micro level	N/A
Web search engine (Google, Seznam etc.)	Low priority	Micro level	N/A

Source: Own research

2.1 Research of the potential of data sources in tourism

In order to determine the potential of new statistical sources, a questionnaire survey of respondents from the experts in the field of tourism is used. The research is mainly focused on the needs of destination management organizations in the Czech Republic. The thematic research focuses on geolocation data, data from payment cards, social networks, shared economy platforms, internet search engines and their comparison with traditional sources. The questionnaire survey was conducted between February and March 2018. 79 tourism experts were questioned. Respondents were geographically distributed within the territory of the Czech Republic. The questionnaire survey was divided into two basic parts. The first part of the research focused on comparing the potential of traditional and new data sources for tourism statistics. The second part of the

MARKÉTA VOGELOVÁ - RADEK CHALOUPKA - JOSEF ABRAHÁM

MARKÉTA VOGELOVÁ - RADEK CHALOUPKA - JOSEF ABRAHÁM

research was devoted exclusively to new data sources of tourism statistics and the way they are used by the tourism stakeholders and the role of the public sector.

The results of the questionnaire survey are divided into 7 specific areas as follows: 1) evaluation of the importance of selected data sources in terms of the needs of respondents to research in the field of tourism statistics, 2) respondents ' confidence in relation to individual data sources as sources for tourism statistics, 3) reasons to prevent respondents from using individual data sources for statistical purposes in the tourism sector, 4) evaluation of future possibilities and potential use of data sources for tourism statistics, 5) preferences of respondents way of providing statistical data from new data sources for the needs of tourism statistics, 6) respondents ' opinion on the payment of statistical data from new data sources a 7) preferences of respondents to regional breakdown statistical data from new data sources and their provision for public use.

3   CONCLUSION

According to the results of the research, experts trust statistical data in tourism, especially those from traditional administrative sources. The highest level of trust have data that experts know and have real experience with them. Data credibility with the level of experience with a given data source is decreasing. The logical result is therefore somewhat lower trust to new sources of tourism statistics. Nevertheless, the results achieved in relation to these resources can be considered favorable.

The questionnaire survey also confirmed the weaknesses of traditional tourism statistics, such as the unpredictability of the data provided, the limited number of statistical indicators and their insufficient regional breakdown. In relation to new data sources, experts see the complexity of the data source methodology and the lack of clarity and level of interpretation of their results as deficiencies. New data sources have been assigned the greatest potential of use, in particular geolocation (signaling) data and Internet search engine services. On the other hand, data about the shared economy are the least potentially usable. Traditional data from the administrative sources of the Czech Statistical Office, the Ministry of Regional Development, the Czech National Bank, and the CzechTourism Agency are also still considered to be promising. In the case of traditional sources, there is still a need to continue in their developing.

The results of the research also brought important impulse for the public sector (especially for the Ministry of Regional Development of the Czech Republic and the Czech Tourism Authority – CzechTourism). According to the results of the questionnaire survey, the public sector should focus mainly on the preparation of a uniform methodology for the use of new data sources and to ensure their full utilization. From the point of view of ensuring full data utilization including uniform methodological unification, geolocation / signaling data and payment

card data are the most preferred. Concerning data from social networks, the public sector should engage rather in methodological support. Very important is also the information on what level the management of tourism should be publicly available or free of charge, and this information is particularly relevant in relation to the newly created destination management structure, which is now established by Categorization of Destination Management Organizations. The survey found that data from new sources should be available publicly and free of charge at national and regional level, with each new data source being different. With a lower level of tourism management, experts' preferences to data provision are declining. Positive is that there is a willingness to pay a reasonable amount of money. Fully available free-of-charge data are especially preferred at national county level. The high interest in obtaining the new data for free is also demonstrated at the regional DMO level.

REFERENCES

AHAS, R., TIRU, M. Passive Anonymous Mobile Positioning Data for Tourism Statistics. In: 11th Global Forum on Tourism Statistics, Reykjavik, 2012.

ARMOOGUM, J., MARCHAL, P., ROUX, S. Acceptability of the use of new technologies by interviewees in surveys. In: New techniques and technologies for tourism, Luxembourg, 2009.

DUPEYRAS, A., MACCALUMM, N. Indicators for Measuring Competitiveness in Tourism. Paris: Organisation for economic cooperation and development, 2013.

EUROSTAT. Tourism Sattelite Accounts in Europe. Luxembourg: Publications Office of the European Union, 2016.

HERGET, J., VOGELOVÁ, M. Monitoring of the project – European Capital of Culture 2015. In: Čechurová, L. (ed.). Opportunities and Threats to Current Business. Management in Cross-border Comparison 2015. Chemnitz: Verlag der GUC, 2015. ISBN 978-3-86367-038-2.

KAMENICKÝ, J., KUČERA, L. Vybrané aspekty cestovního ruchu České republiky. Praha: Český statistický úřad, 2014.

SEYNAEVE, G., DEMUNTER, Ch. When mobile network operators and statistical offices meet - integrating mobile positioning data into the production process of tourism statistics. In: 14th Global Forum on Tourism Statistics, Venice 2016.

WORLD ECONOMIC FORUM (WEF). The Travel and Tourism Competitiveness Report 2017. Geneva: WEF, 2017.

WORLD TOURISM ORGANIZATION. UNWTO Tourism Highlights: 2018 Edition. Madrid: UNWTO, 2018.

WORLD TRAVEL & TOURISM COUNCIL (WTTC). Regional results (World). London: WTTC, 2018.