sciendo

Big Data: a Source of Mobility Behaviour and a Strategic Tool for Destination Management

Filip Emmer / Andrea Holešinská
e-mail: holesinska@econ.muni.cz
Institute of Tourism, Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic

Emmer, F., & Holešinská, A. (2020). Big Data: a Source of Mobility Behaviour and a Strategic Tool for Destination Management. *Czech Journal of Tourism,* 8(2), 85–102. DOI: 10.2478/cjot-2019–0006.

Abstract

The abundant use of the Internet and mobile technologies while traveling leaves a digital footprint in the form of big data that can be tracked. Big data bring information about spatial visitor behaviour that is valuable for strategic destination management. Big data enrich not only scientific fields (e. g. management, marketing, or geography) with their knowledge, but also represent the invention of new tools for their actual processing. Generally, big data are considered as a strategic tool enhancing the competitiveness of a destination. The paper presents the basic characteristics of big data and reviews research focused on big data in tourism. Moreover, it identifies its potential for tourism from both the theoretical and methodological point of view. The final part deals with current trends in using the big data in tourism and its application in destination management. The future trends of big data in the context of destination management are implied as well.

Keywords

tourism, big data, mobility behaviour, destination management

JEL classification: C81, M15, Z32

Introduction

The long-term economic growth of most countries in the world and the associated trend of increasing population wealth have enabled the development of mass tourism, in which hundreds of millions of people participate every year. In combination with the massive development of information and communication technologies (ICT) as described by Kambatla, Kollias, Kumar, and Grama (2014), there have been significant changes in the behaviour of tourism entities, not just the Business to customer segment, but also Business to business and customer to customer (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011). The emergence of the Internet, which has been evident since the beginning of the new millennium, has created a platform that facilitated communication and accelerated a number of tourism-related activities (Li & Wang, 2011; Thevenot, 2007). A good example is the recently created social networks that allow sharing of photo experiences, as well as opinions and feedback, all in near real time. By their importance, Xiang and Gretzel (2010) identify them as one of two megatrends having a strong impact on tourism (the other being search engines). Support and information services are equally important for tourism. Today, it is no longer necessary to leave the comfort of your home or office when planning a vacation, travel agencies are not required to get the best prices, and a person no longer has to rely on oral recommendations when booking a hotel. Everything can be found in just a few minutes thanks to the Internet on the website, whether destinations or comparison sites. Such a way of searching, interacting, and subsequent decision-making processes leaves a considerable digital footprint that provides a significant amount of data. Heerschap, Ortega, Priem, and Offermans (2014) speak of the transition from "data scarcity to data abundance" to the growing volume of data. Kambatla, Kollias Kumar, and Grama (2014) point to the volume of processed information, which in the private business sector alone amounted to 9.57×10^{21} bytes of data, and which was estimated to double every two years. It is this, if not the only, characteristic that gave such data a name. These are big data.

Information is of great value on the market. They represent a competitive advantage. Today, the source of information is big data that are generated by the fact that "customers leave electronic traces during all travel-related activities" (Fuchs, Höpken, & Lexhagen, 2014). This creates data on spatial tourist behaviour (i.e. mobility and tourist decision-making process), which are important for the strategic management of the destination, see Baggio and Scaglione (2018).

The aim of this paper is to identify the existing trends in the use of big data with an emphasis on their use in the destination management area and to outline the possible future trends in the use of big data for the purpose of destination management.

The questions are as follows:

- Object-related, i.e. What are big data? What are their characteristics? What is their source/ Where do they come from?
- Subject-related, i.e. How can big data be used in tourism? Who can benefit from the results? How can big data help enhance the competitiveness of a destination?

- Methodology-related, i.e. How to process big data? What is the contribution of big data? What are the limitations of big data?

Big data

The concept of big data may appear to be a relatively new term, unknown to the general public. However, the essence of big data has been perceived indirectly since the origin of the first computers and their generated outputs (Press, 2013). In the coming years, the increasing capacity of the generated data began to be reflected in the effort of both scientists and commercial companies, increasingly storing this data electronically, for example, in the form of powerful compressors (Marron & Maine, 1967). At the beginning of the millennium, it was clear that entities with large volumes of data faced new challenges. Laney (2001) emphasized in particular the role of data management. However, as Snijders, Matzat, and Reips (2012) point out, big data are defined very vaguely, most often as "datasets so large and complex that they become awkward to work with standard statistical software". A narrower specification is offered by Laney (2001), who characterized the properties of these emerging data sets and summarized them into the 3V concept (volume, velocity, variety), which is still the starting point for defining big data. In the follow-up studies, several modifications of this concept appear, for example, the 7V concept (Yu, Choi, Shin, & Ahn, 2014). For the purpose of this text, it is essential to define big data using the 3V concept and selected sub-characteristics.

Big data characteristics

Volume

The basic feature of big data is its volume. It is created by the continuous generation of data, mainly through various platforms operating on the Internet - the so-called Internet of Things (Botta, de Donato, Persico, & Pescapé, 2016), which nowadays interconnects an increasing number of devices of everyday life and thus forms a connected network, through which it exponentially increases the data obtained (Gubbi, Buya, Marusic, & Palaniswani, 2013). The volumes of the data generated are doubled every two years and their capacity is estimated to reach 44 zettabytes, equivalent to 1 billion terabytes (EMC, 2014). Such an increase in data was predicted at the beginning of the millennium, as data storage costs gradually decreased, while communication through electronic channels was made easier and cheaper, thus reducing transaction costs for data forwarding as such (Laney, 2001). The creation of increasingly large amounts of data is also related to the behaviour of businesses that have identified the importance of information collected from the data that can significantly affect their market behaviour and maximize profit, leading to ongoing data retention, including legacy data. However, this long-term trend of data retention shows a declining marginal benefit from each additional data point stored (Laney, 2001). As the marginal utility decreases, the so-called Digital Universe Paradox can be observed. Although the cost of the data

storage and transmission has fallen sharply over time, companies are investing more and more in data management tools.

In the current research of big data in the field of tourism one can observe the irreplaceable role of this factor. Vu, Li, Law, and Ye (2015) work with a sample of 29,443 photos from Hong Kong, Phillander and Zhong (2016) in its sentimental analysis based on a dataset of 31,550 samples covering more than 30 hotels, Olmedo, Goméz, Palomares, and Gutiérrez (2018) are based on a comparison of big data sources from a sample of 307,062 geo-located photos and 234,159 tweets from Twitter related to landmarks in Madrid.

The challenge for data management, which destination management has to cope with limited resources and ever-increasing volumes, may be another characteristic of big data. This second V is variety. The very nature of big data, consisting of a large volume, also makes it impossible to use some standard or classical statistical methods for calculations, which can make it more difficult to use them.

Variety

The second characteristic of big data can be understood in several levels. The first is the variety of big data sources. This article is based on the distribution reported by Li, Xu, Tang, Wang, and Li (2018), which differentiates data according to the following three sources of their origin. The first are the data generated by users themselves (typically Twitter reviews, blogs, Instagram photos), otherwise referred to as user-generated data (UGC data), data generated by devices. For the purpose of tourism, the important are especially the data from mobile phones or sim cards, data from navigation and other devices using GPS data and transaction data, whether on-line booking, credit or debit card payments and cash withdrawals, all of which are the source of the so-called transaction data (Xiang, Du, Ma, & Fan, 2017). The second level of perception of the variety of the data obtained is based on their sources; it is the very form in which the data is organized. The UGC data have probably the greatest variety from these sources. Photographs, which offer a source of information both from the place of acquisition as well as from their description or information about the author, can be processed, reviews and tourists' evaluation have a textual form, which is usually subjected to a sentimental or lexical analysis. A lot of useful information and data are also increasingly obtained from Youtube videos.

However, the variety of data also entails a variety of difficulties in obtaining it. East, Osborne, Kemp and Woodfine (2017) used GPS locators to map the movement of visitors to the Marwell Zoo in the UK. They state that "the accuracy of the GPS units was affected by the buildings that the participants entered and by the dense tree cover. Accuracy in the more open areas of the park was typically ± 5 m directly adjacent to the woodland accuracy in the region of ± 8e10 m". Despite these and other problems with GPS data, research has made it possible to identify very precisely the density of the visitor stream, which also varied over time. As the authors mention, "This bunching and

peaking in visitor numbers can be of benefit to the institution as it allows visitor presentations to be timed to coincide with the peak in visitor numbers at a given location". Philander and Zhong (2016) as one of the main limits of social network data (specifically Twitter) report that their "Analysis is limited to the extent that social media platform owners are willing to share their data, or that data can be legally mined". At the same time, "Twitter API provides limited search capacity regarding returned data and time frame". Such output cannot be reliably provided by a classical questionnaire survey. The vast majority of this information and data is generated in hours and sometimes minutes, thus defining the third basic characteristic of big data, which is velocity.

Velocity

The velocity of data primarily refers to the rate at which data are generated. This velocity is determined by the source platform, i.e. online data sources and sensors, but for some types of data also by the immediacy of their profit and the absence of interaction with the people who directly produce the data. This principle can be presented on the difference of collecting data about the movement of tourists in the destination, for which the data from mobile phones can be used, but the same problem can be investigated by a questionnaire survey. While a significant amount of time is needed to collect data from individual respondents in terms of data generation velocity, the data from mobile phones, GPS and sim cards are generated almost immediately. Heerschap, Ortega Priem, and Offermans (2014) point out the advantages of mobile data more specifically: "It is possible to, for example, zoom in on small specific tourist areas and very short time frames, such as certain events, coastal areas, or national holidays; to distinguish data per country of residence; or determine the number of transit passengers. Kitchin (2013) highlights the automatically generated data and its velocity:" Such systems include immigration passport control where passenger details are collected and checked against various databases in real-time, and new data are generated such as CCTV, photographs, fingerprints or iris scans". Especially the real-time data creation and processing are often crucial for economic operators. The ability to respond flexibly to market developments makes it possible to better exploit opportunities or respond to the changes in demand trends (Che, Safran, & Peng, 2013).

Veracity

The findings of research studies indicate the "uncertainty" of the truth of the output from the big data analysis (Lukoianova & Rubin, 2014). This uncertainty is determined by the origin, analytical tools and methodology used to process big data (Lukoianova & Rubin, 2014). The reliability of the results may also be distorted by the incompleteness of the data in terms of population representation, as the demographically older part of the population may not be sufficiently represented, if at all, in the data. Vaynerchuk (2013) points out that big data obtained from social networks may not be complete because "twitter users skew towards younger and urban groups".

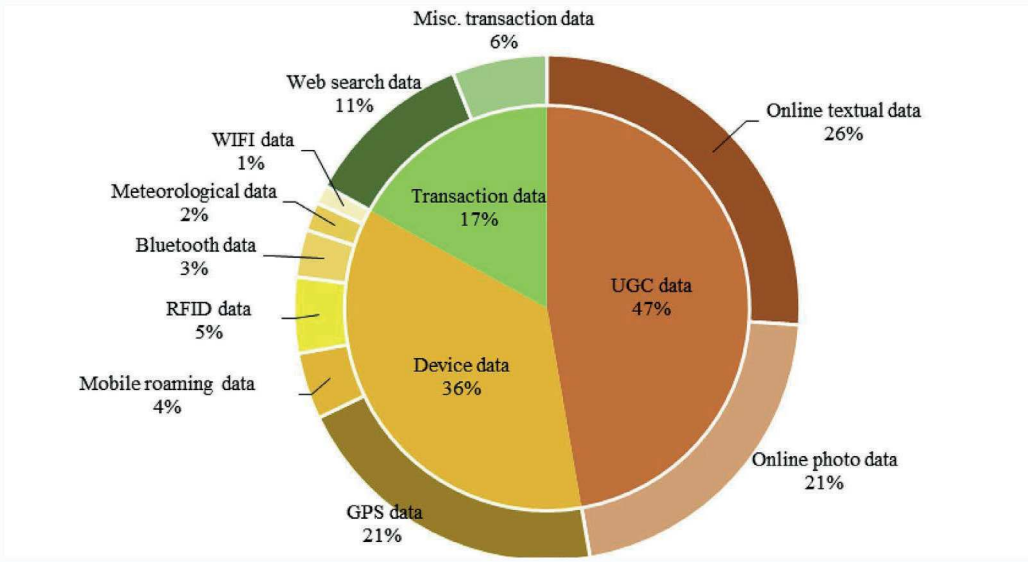
Vulnerability

Many big data sources are based on personal information, typically the data from mobile applications, sim cards and payment cards. There have been a number of cyber-attacks in the recent times aimed at stealing email addresses and accounts on Linkedln (The Guardian, 2012). The issue of the security of such data appears to be one of the key issues in tourism, given the recent events, such as the terrorist attacks in France and Belgium, which could otherwise be targeted with more serious consequences if sensitive data were used in the future. A double-edged weapon may also be the ignorance of tourists themselves about the collection and evaluation of these data. The existing data is "most of the time based on information recalled by the interviewees" (Baggio & Scaglione, 2017), allowing the respondent the right not to answer. On the other hand, some forms of big data are obtained entirely independently of tourist awareness and, although the data are anonymised in the vast majority of cases, they potentially pose a sensitive personal and security risk.

Big data resources

In relation to the variety of big data, a diversity of potential data sources is essential for destination management as well as research on big data. Li, Xu, Tang, Wang, and Li (2018) divide them into three categories, the data generated by the user himself, the data automatically generated by machines or chips, and finally, the data generated by the transactions (consult Figure 1 below).

Figure 1 Big data sources used for tourism research purposes and their representation



Source: Li, Xu, Tang, Wang, & Li (2018)

UGC data

The user-generated content (UGC) data have gained in importance due to the boom of social networks. These enable content to be distributed at high speed and with wide reach. This source can be divided into the online text data, which consist mainly of opinions and reviews, and the online photo data.

On-line text data

Opinions and reviews of tourism participants form a valuable source for quality management, identification of opportunities in terms of implementation prerequisites, but also in identifying the key stakeholders in the destination. Research on the online text data is conducted by Olmedo, Gómez, Palomares, and Gutiérrez (2018), who compare different sources of big data in tracking and monitoring the activities of Madrid travellers, one of which is geo-localized social networking on Twitter. This social network is suitable for a data collection and analysis because of its relative openness in providing access to APIs (application programming interfaces). This is also used by Phillander and Zhong (2016), who, using a sentimental analysis, quantify and evaluate the customer experience of "tweeting" tourists staying in hotel establishments in Las Vegas. The benefit of this method is that the sentimental analysis "can provide a quantifiable measure of real-time marketing campaign impacts complementary to other traditional research methods". Ye, Law, Gu, and Chen (2011) reviewed the impact of reviews on travel sales to a destination. By analysing contributions to the Chinese reservation portal ctrip.com, where only those customers who have made and paid an order can write their experience of quality of service.

On-line photo data

With massive web development 2.0 and the emergence of online photo albums, blogs and social networks, the reach and potential audience that can see these photos has also increased (Boyd, 2007). This audience then also creates an image, preferences and motivation about the destination to visit or not to visit. Many potential tourists even consider such sources of information more relevant than the officially published documents and leaflets that are usually used by the Tourism Headquarters and DMOs (Mack, Blose, & Pan, 2008; Xiang & Gretzel, 2010). Currently, the most used platform for big data is Flickr. Flickr is a video and photo hosting service that has over 75 million active users worldwide (www.flickr.com/jobs) who upload over 1.6 million photos daily (www.flickr.com/photos). Lo, McKercher, Lo, Cheung, and Law (2011) have shown that 89% of Hong Kong residents take photos on their recreational trips, with 41% sharing them online and 40% posting them directly to Flickr or a similar website. In research, however, they encounter one of the typical big data problems when they point out that "At present, a large demographic gap exists in both the profile and importance of travel between those who do and do not post their photographs online. The gap will likely become more acute as online photo-sharing media continue to gain in popularity with a larger share of the population." The Flickr data allow to accurately locate where the

tourist was. If s/he uploaded more than one photo, it is possible to determine what her/his movement in the destination looked like based on time delays and geolocation. Kurashima, Iwata, Irie, and Fujimura (2010), for example, have developed a system of referrals for places to visit, based only on tracking and evaluating photos from Flickr. This system calculates the likelihood of a site visit, taking into account the user's preferences and current location. Olmedo, Gómez, Palomares, and Gutiérrez (2018) use another Panoramio photo platform in their comparison of big data sources, but this is no longer available today. Instagram, currently the largest photo and video sharing platform, also embodies high potential, but it provides almost no data files for its security policy.

Device data

This kind of data originated with the massive development of the so-called Internet of Things, i.e. the devices that are constantly connected to the Internet. In combination with many types of sensors and sensors available to these devices, the devices belonging to the Internet of Things category offer potential, especially in mapping movement and spatial habits. The main technologies used to make such data types include GPS, bluetooth and WIFI, as well as the RFID technologies, working on the principle of radio frequency waves (Derakhshan, Orlowska, & Li, 2007).

GPS data

These data are significant compared to the previous types, especially the number of users who generate them, knowingly or unknowingly, and because they are obtained directly in the destination and given space. In particular, two approaches are used for the collection of such data types. The first is the data collection directly from mobile phone applications that use location services. This approach is particularly suitable for larger areas and for analysing a larger number of tourists. The second approach mentioned by Orellana, Bregt, Ligtenberg, and Wachowicz (2012) is the distribution of GPS tracking devices, for example, at the entrances of fun parks or monuments. Using these monitoring devices, the movements of tourists in the area are subsequently monitored and evaluated. Zheng, Huang, and Li (2017) use the GPS technology to collect and evaluate data in the Beijing Summer Palace area. These data are used to predict spatial behaviour and distribution of tourists in the destination. Tchetchik, Fleischer, and Shoval (2009) use the GPS data in combination with a questionnaire survey to segment tourists in the historic Acre area.

Mobile data

Similar to the GPS data, information about the movement of sim cards and mobile devices allows very precise observation of the movement of tourists. Compared to the GPS data, however, they can originate in two, very different ways. These data are divided into two types. The first is the so-called Log data, i.e. the data obtained directly from the subject's mobile phone through an application installed on the device. At 5-minute

intervals, such an application sends information about the location and time and all the operations that the mobile phone has performed during this time period (Heerschap, Ortega, Priem, & Offermans, 2014). The second type of mobile data are the anonymous data based on call details (Call detail records). These data are generated by telecommunications service providers. Every time a mobile device contacts an operator (SMS, phone, or data download), a record of the location and time where the interaction took place is created. The applicability of the analysed data has been demonstrated by Heerschap, Ortega, Priem, and Offermans (2014), while highlighting the limits and challenges for further development, for example, "results of measurements based on the big data sources also require new methods of data visualisation, such as geo-location maps with time indicators. The challenge here is to produce visualisations that can be clearly underwood and interpreted by users, such as policy-makers."

Bluetooth data

A wide subset of mobile data are the data related to signal transmission. These include Wifi data and Bluetooth data. The visitor behaviour in Ghent using the bluetooth data was examined by Versichele, de Groote, Bouuaert, Neutens, Moerman, and Van de Weghe (2014).

Transaction data

The transaction data can be divided into the data used for online payment and search strategies and the data generated by using credit and gift cards and payments directly at the destination. As reported by Jansen and Molina (2006), a search query analysis (the case of the first data source) makes it possible to determine the user's strategy for obtaining information, typically search intent, search depth, or keywords used by the user. Wöber (2007) divided the metropolis into two types by analysing searches related to European cities. The first group (Madrid, Budapest, Prague, Nice, or Rotterdam) was dominated by searches for guided tours, operas, museums, and art. While in the second group (Heidelberg, Talinn, Copenhagen, Lyon), questions related to the attractions and events prevailed. Similar data sources include the data from reservation systems used by Saito, Takahashi, and Tsuda (2016) to analyse customer behaviour based on the data obtained from the reservation system. The default dates are the booking data for four hotels in Kyoto. The data include the number of bookings, the prices of individual bookings and their booking dates. The results of this research were used in the following years to predict the long-term development of consumer behaviour. In contrast, Shih, Nicholls, and Holecek (2009), for example, used data from eight consecutive years, based on sales of ski passports in the Michigan area (physical sales data) to examine the effect of weather changes on sales.

On the contrary, Weaver (2008) demonstrates usefulness of loyalty and gift cards: "The portability of these cards makes them a convenient companion to mobile consumers. For example, reward cards enable a hotel chain to collect data from consumers even

if they stay at a number of its properties across an entire country or in several countries. Preferences expressed by hotel customers can be recorded and stored within a centralized database."

Big data in tourism

Baggio and Scaglione (2018) clearly describe a shift from the traditional data to big data. At the beginning, there were traditional methods of data collection (surveys or opinion polls). Baggio (2016) called them small data approaches. With the development of social media, the data generated via geo-tagged photos, the so-called volunteered geographical information started to be widely used (Baggio & Scaglione, 2018). Another data source that has increased its importance due to business intelligence (Höpken & Fuchs, 2016), are the transaction data linked to destination guest cards. Many researchers use the data generated by global positioning systems (GPS). Nowadays, a frequently used data source include passive mobile phone positioning data.

Table 1 Big data sources and research cases

Data source categories by Li, Xu, Tang, Wang, and Li (2018)	Data source categories by Baggio and Scaglione (2018)	Cases
-	Small data approaches (surveys/opinion polls)	Hwang, Gretzel, and Fesenmaier (2006) Fuchs, Höpken, and Lexhagen (2014) Haarshap, Ortega, Priem, and Offermans (2014) Sabou, Onder, Brasoveanu, and Scharl (2016)
User-generated content data — On-line photo data	Volunteered geographical information (geo-tagged photos)	Zach and Gretzel (2011) Kádár and Gede (2013) Vu, Li, Law, and Ye (2015) Stienmetz and Fesenmaier (2016)
— On-line text data		Philander and Zhong (2014) Ye, Law, Gu, and Chen (2011)
Transaction data	Destination guest cards	Fuch, Abadzhiev, Svensson, Höpken, and Lexhagen (2013) Höpken, Fuchs, and Lexhagen (2013) Zoltan and McKercher (2015) Holland, Jacobs, and Klein (2016)

Device data GPS data	GPS	Shovel and Isaacson (2007) Zheng, Huang, and Li (2017) East, Osborne, Kemp, and Woodfine (2017)
Mobile data	Passive mobile phone positioning data	Haarschap, Orlega, Priem, and Offermans (2014) Baggio and Scaglione (2018) Saluveer, Ruan, Tiru, Altin, Kroon, Snitsarenko, Aasa, and Silm (2020)

Source: own processing

Big data offer potentially great benefits. Höpken, Ernesti, Fuchs, Kronenberg, and Lexhagen (2017) see it in the prediction of arrivals. Therefore, it is necessary to examine the spatial behaviour of tourists (Olmedo, Gómez, Palomares, & Gutiérrez, 2018). It is possible to avoid (overtourism), a large concentration of demand to a few points of interest in the destination. Big data also allow the issue of shared housing, which has significant socio-economic impacts on residents in the destination (Zheng, Huang, & Li, 2017).

Baggio and Scaglione (2018) highlight the use of big data to interpret visitor flows and use this information for strategic destination management. Destination management information systems (DMIS) are important for effective planning, leading to a reduction in information asymmetry among tourism stakeholders (Fuchs, Abadzhiev, Svensson, Höpken, & Lexhagen, 2013). Knowledge of information about customer behaviour thus contributes to the competitiveness of the destination. Big data are important for destination management organizations (DMOs), but also for tourism stakeholders from both private and public sectors as they enter DMIS. According to Fuchs, Abadzhiev, Svensson, Höpken, and Lexhagen (2013) "individual stakeholders may use data for improving their own activities somewhat independently of others, their relationship is more interdependent at the strategic destination level, what calls for more integrated and coordinated knowledge application processes". Baggio and Scaglione (2017) point out that "from the point of governance, ... the mobile data network analysis would be perceived by stakeholders as more objective than experts opinion panels".

The benefit of using big data is undoubtedly knowledge about the visitor's mobility and decision-making behaviour, and thus the possibility to better influence or (adapt) to her/his needs, i.e. to better (more effectively) target and compose the offer, and at the same time the possibility to affect the visitor's behaviour in the destination with respect to the use of recreational resources. Thus, the benefits can be seen both in terms of competitiveness (access to information and more detailed visitor knowledge) and sustainability (more targeted visitor management), see the model of the competitive destination Ritchie and Crouch (2006). In general, big data have the potential to streamline the planning process.

Big data and methodology

Most authors (e.g. Philander & Zhong, 2016; Heerschap, Ortega, Priem, & Offermans, 2014) agree that big data make it possible to improve or refine classic statistical outputs, enriching them with information that can be specific to DMO and decision-making processes, such as information on consumer preferences, more detailed tourist flows, or potential stakeholders based on user reviews.

This knowledge contributes to strategic planning as a part of the management mechanism. The implication of big data extends the background of knowledge-based destination, business intelligence, organizational learning (networks), destination management (mobility behaviour), visitor management (spatial distribution), or marketing (decision-making process). For instance, Beritelli (2019) uses big data for the description of visitor flows and consequently, for the inference of the lifecycle of destination.

From the methodology point of view, the review of quoted studies about big data shows that the most frequently used research methods for data processing are coding, classification, and clustering (see Table 2). It is mostly processing of the integrated (both qualitative and quantitative) data from different sources, e.g. external data (web content), unstructured data (customer reviews), and data in real-time. Therefore, it is necessary to use research tools such as data mining software.

Table 2 An overview of frequently used tools and methods

Author	Tools	Research methods	Object of research
Önder (2017)		Classification	Multi-destination trips
Sabou, Onder, Brasoveanu, and Scharl (2016)	ETIHQ Dashboard (data analytics system) semantic web technologies	ontologies applied encoding	organizational learning at tourism destinations
Fuchs, Abadzhiev, Svensson, Höpken, and Lexhagen (2013)	business intelligence system - data mining software (DMIS)	Classification, Clustering	Customer behaviour in all trip phases (new knowledge)
Francalanci and Hussain (2015)		Clustering, Network analysis (visualization)	multi-layered clusters of nodes surrounding hub nodes
Sun, Wei, Tsui, and Wang (2019)	Search query data; Kernel-based extreme learning machines	Indexing, Generalization,	Forecast tourism demand (arrivals)

Source: own processing

No matter how rich big data are, there are certain limits. Big data cannot replace classic statistical sources and cannot provide results that would satisfy DMO in broad-spectrum decision-making in the light of the classic DMO roles (Philander & Zhong,

2016; Heerschap, Ortega, Priem, & Offermans, 2014). One of the reasons is the cost of obtaining these data and the amount of big data sources that would be necessary for such a comprehensive analysis.

Phillander and Zhong (2016) describe other limits of the use of big data. One of them is a sample of the population that generates the data. Although modern technologies have spread across the whole of society, their normal use is not evenly distributed across the age spectrum. An obvious example is those in the old-age pensioner category who travel frequently, but their footprint may not be recorded at all. Another important problem mentioned by these authors is the ethical problem since big data allow the collection of highly sensitive data that could be misused very easily in case of leakage.

Conclusion

Using big data to visualize visitor behaviour in a destination is crucial for the strategic development of the destination. They provide information about visitor mobility within the destination and at the same time identify the most visited sites.

Probably the most widespread source of big data are the Mobile Positioning Data, which enable to obtain better cross-border statistics on the movement of tourists. Their main asset, however, is the excellent coverage in the collection of data on multi-day stays, as they make it possible to describe 'non-registered' or 'non-paid' overnight stays. This can also be used to map the Airbnb phenomenon and, in combination with the data on housing price developments, this information can serve to better target measures and possibly regulate the phenomenon of shared housing.

The potential of big data may conceal the facilitation of solving many problems in relation to the supply and demand of tourism, but it should be considered that "progress and discoveries eliminate many obstacles, but at the same time put new obstacles and threats" (Bastiat, 2015). In the case of big data, this is particularly the question of privacy and the potential misuse of this, often very sensitive, information.

The massive use of big data in the IT industry has only begun in recent years. For the time being, these alternative data sources are used only marginally in tourism. There is still a strong dependence on the collection and evaluation of data from national or international institutions. However, this does not mean that current research neglects this area, on the contrary. As this article shows, the medium-term trend of exploring big data is growing.

Although the use of big data for destination management has been researched only by a small number of authors, the potential in this area is quite evident. The variety of big data, which is one of the basic characteristics of these data, allows the data to form a kind of complementary relationship with one data source to quantify a certain phenomenon in a destination (for example, travel around tourist attractions) and another source enables to explore the qualitative aspects of the same (for instance, feelings of destination through the sentimental analysis of social data). Such a comprehensive picture of all activities in the destination would give destination management organisations

(DMOs) a whole new dimension of interaction with stakeholders in the destination in the context of the 3C (cooperation-coordination-communication) principle (Holešinská, 2012).

Although big data is of great value for regional and local DMOs because they enable a high level of accuracy of information, at present big data are used rather at macro-regional level. This is due to the lack of staffing and financial capacity required for data mining or data purchasing and analysis. In such an environment, it is therefore natural that there is a trend to use big data as a refinement of one particular monitored phenomenon in the destination.

The future development of big data use can be illustrated by the example of the portal of the Swiss national project Grandtour (https://grandtour.myswitzerland.com/en/), which among other things allows users to share experiences and connects them on social networks with a widely used hashtag. The set of photos and their descriptions marked with this hashtag allows to analyse the places visited, the visitor's opinion, their emotions, but also the extent that such a contribution may have.

From the economic point of view, however, the current trend of using big data is evident. Due to the high purchase price compared to other data sources, big data are nowadays a luxury property for many stakeholders in tourism. But what is the luxury of today, is the necessity of tomorrow (Von Mises, 1998). Big data were generated as a result of advances in the field of profit, processing and evaluation of data from the most modern devices using the internet, mobile or other kind of signal and it will take some time, if any, to establish themselves in the tourism environment. However, if their value is proven, they can be widely used.

In this respect, tourism policy should be involved by supporting digitization activities while creating the right conditions for using big data for strategic destination management. For DMO, working with big data in the future is crucial, and even strategic.

Acknowledgment

This article was supported by the project New Mobility – High-Speed Transport Systems and Transport-Related Human Behaviour, reg. number CZ.02.1.01/0.0/0.0/16_026/0008430, code for CEP (EF16_026/0008430) at Masaryk University in Brno.

References

Baggio, R., & Scaglione, M. (2017). Strategic Visitor Flows (SVF) Analysis Using Mobile Data. In Schegg, R., Stangl, B. (Eds.), *Information and Communication Technologies in Tourism 2017* (pp. 145–157). Cham: Springer. DOI: 10.1007/978-3-319-51168-9_11.

Baggio, R., & Scaglione, M. (2018). Strategic visitor flows and destination management organization. *Information Technology & Tourism*, 18, 29–42. DOI: 10.1007/s40558-017-0096-1.

Baggio, R. (2016). Big data, business intelligence and tourism: a brief analysis of the literature. In Fuchs, M., Lexhagen, M., Höpken, W. (Eds.), *IFITT workshop on the big data and business*

intelligence in the travel and tourism domain. European Tourism Research Institute (ETOUR). Östersund: Mid-Sweden University.

Bastiat, C. F. (2015). *Petice za zákaz slunce a jiné absurdity ekonomie* [Petitions for the prohibition of the sun and other absurdities of economics]. Prague, Czech Republic: Mises.cz.

Beritelli, P. (2019). Transferring concepts and tools from other fields to the tourist destination: A critical viewpoint focusing on the lifecycle concept. *Journal of Destination Marketing & Management*, 14. DOI: 10.1016/j.jdmm.2019.100384.

Botta, A., de Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future Generation Computer System*, 56(C), 684–700. DOI: 10.1016/j.future.2015.09.021.

Boyd, D. (2007). Why Youth Social Network Sites: The Role of Networked Publics in Teenage Social Life. In Buckingham, D. (Ed.), *Youth, Identity, and Digital Media, Beckman Center Research Publication No. 2007–16*. Cambridge, USA: The MIT Press. Retrieved from https://ssrn.com/abstract=1518924.

Che, D., Safran, M., & Peng, Z. (2013). From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. In Hong, B., Meng, X., Chen, L., Winiwarter, W., Song, W. (Eds.), *DASFAA 2013: Database Systems for Advanced Applications,* vol. 7827 (pp. 1–15). Verlag/Berlin/Heidelberg: Springer. DOI: 10.1007/978-3-642-40270-8_1.

Derakhshan, R., Orlowska, M. E., & Li, X. (2007). RFID Data Management: Challenges and Opportunities. In *2007 IEEE International Conference on RFID, March 26–28* (pp. 175–182). Grapevine, USA: IEEE. DOI: 10.1109/RFID.2007.346166.

East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, 61, 307–320. DOI: 10.1016/j.tourman.2017.02.021.

EMC Digital Universe with Research & Analysis. (2014). Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives. *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things* [online]. Retrieved from https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm.

Francalanci, C., & Hussain, A. (2015). Discovering social influencers with network visualization: evidence from the tourism domain. *Information and Communication Technologies in Tourism*, 16, 103–125. DOI: 10.1007/s40558-015-0030-3.

Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework for tourism sustainability: A business intelligence application from Sweden. *Tourism*, 61(2), 121–148.

Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management*, 3, 198–209. DOI: 10.1016/j.jdmm.2014.08.002.

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. DOI: 10.1016/j.future.2013.01.010.

Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014). Innovation of tourism statistics through the use of new big data sources. In 12*th World Telecommunication/ICT Indicators Symposium (WTIS-14), Tbilisi, Georgia, 24–26 November 2014*. ITU: Geneva, Switzerland. 12p. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/events/wtis2014/002INF-E.pdf.

Holešinská, A. (2012). *Destinační management jako nástroj regionální politiky cestovního ruchu* (Destination management as a tool for regional tourism policy). Brno, Czech Republic: Masaryk University.

Holland, C. P., Jacobs, J. A., & Klein, S. (2016). The role and impact of comparison websites on the consumer search process in the US and German airline markets. *Information Technology and Communication in Tourism*, 16, 127–148. DOI: 10.1007/s40558–015–0037–9.

Höpken W., Ernesti D., Fuchs M., Kronenberg K., & Lexhagen M. (2017). Big Data as Input for Predicting Tourist Arrivals. In Schegg R., Stangl B. (Eds.), *Information and Communication Technologies in Tourism 2017* (pp. 187–199). Cham: Springer. DOI: 10.1007/978–3-319–51168–9_14.

Höpken, W., & Fuchs (2016). Introduction: Special Issue on Business intelligence and big data in the travel and tourism domain. *Information and Communication Technologies in Tourism*, 16, 1–4. DOI: 10.1007/s40558–016–0054–3.

Höpken, W., Fuchs, M., & Lexhagen, M. (2013). The knowledge destination – applying methods of business intelligence to tourism. In Wang, J. (Ed.), *Encyclopedia of Business Analytics and Optimization*. Pennsylvania: IGI Global Publisher.

Hwang, Y. H., Gretzel, U., & Fesenmaier, D. R. (2006). Multicity trip patterns: Tourists to the United States. *Annals of Tourism Research*, 33(4), 1057–1078. DOI: 10.1016/j.annals.2006.04.004.

Jansen, B. J., & Molina, P. R. (2006). The effectiveness of Web search engines for retrieving relevant ecommerce links. *Information Processing & Management*, 42, 1075–1098. DOI: 10.1016/j.ipm.2005.09.003.

Kádár, B., & Gede, M. (2013). Where do tourists go? Visualizing and analysing the spatial distribution of geotagged photography. *The International Journal for Geographic Information and Geovisualization*, 48(2), 78–88.

Kambatla, K., Kollias, G., Kumar, V., & Grama. A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561–2573 DOI: 10.1016/j.jpdc.2014.01.003.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241–251. DOI: 10.1016/j.bushor.2011.01.005.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267. DOI: 10.1177/2043820613513388.

Kurashima, T., Iwata, T., Irie, G., & Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. In *CIKM '10: Proceedings of the 19th ACM international konference on Information and knowledge management*, October 26–30 (pp. 579–588). Toronto, Canada: ACM. DOI: 10.1145/1871437.1871513.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. In Application Delivery Strategies, File 949. Stamford, USA: META Group Inc. Retrieved from https://blogs.gartner.com/doug-laney/files/2012/01/ad949–3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68(October), 301–323. DOI: 10.1016/j.tourman.2018.03.009.

Li, X., & Wang, Y. C. (2011). China in the eyes of western travelers as represented in travel blogs. *Journal o Travel & Tourism Marketing*, 28(7), 689–719. DOI: 10.1080/10548408.2011.615245.

Lo, I. S., McKercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism Management*, 32(4), 725–731. DOI: 10.1016/j.tourman.2010.06.001.

Lukoianova, T., & Rubin, V. L. (2014). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances In Classification Research Online*, 1(1), 1–13. DOI: 10.7152/acro.v24i1.14671.

Mack, R. W., Blose, J. E., & Pan, B. (2008). Believe it or not: Credibility of blogs in tourism. *Journal of Vacation Marketing*, 14(2), 133–144. DOI: 10.1177/1356766707087521.

Marron, B. A., & de Maine, P. A. D. (1967). Automatic data compression. *Communications of the ACM*, 10(11), 711–715. DOI: 10.1145/363790.363813.

Olmedo, M. H. S., Gómez, B. M., Palomares, J. C. G., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13–25. DOI: 10.1016/j.tourman.2017.11.001.

Önder, I. (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives*, 21, 54–58. DOI: 10.1016/j.tmp.2016.11.002.

Orellana, D., Bregt, A. K., Ligtenberg, A., & Wachowicz, M. (2012). Exploring visitor movement patterns in natural recreational areas. *Tourism Management*, 33(3), 672–682. DOI: 10.1016/j.tourman.2011.07.010.

Philander, K., & Zhong, Y. Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55, 16–24. DOI: 10.1016/j.ijhm.2016.02.001.

Press, G. (2013). *A Very Short History of Big Data* [online]. Forbes, May. Retrieved from https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#2549c2f865a1.

Ritchie, J. B., & Crouch, G. I. (2006). *The competitive destination: a sustainable tourism perspective*. Wallingford: CABI Publishing.

Sabou, M., Onder, I., Brasoveanu, A. M. P., & Scharl, A. (2016). Towards cross-domain data analytics in tourism: a linked data based approach. *Information and Communication Technologies in Tourism*, 16, 71–101. DOI: 10.1007/s40558–015–0049–5.

Saluveer, E., Raun, J., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., Aasa, A., & Silm, S. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research*, 81, 102895. DOI: 10.1016/j.annals.2020.102895.

Shih, C., Nicholls, S., & Holecek, D. F. (2008). Impact of Weather on Downhill Ski Lift Ticket Sales. *Journal of Travel Research*, 47(3), 359–372. DOI: 10.1177/0047287508321207.

Shoval, N., & Isaacson, M. (2007). Tracking tourists in the digital age. *Annals of Tourism Research*, 34(1), 141–159. DOI: 10.1016/j.annals.2006.07.007.

Siato, T., Takahashi, A., & Tsuda, H. (2016). Optimal room charge and expected sales under discrete choice models with limited capacity. *International Journal of Hospitality Management*, 57, 116–131. DOI: 10.1016/j.ijhm.2016.06.006.

Snijders, C., Matzat, U., & Reips, U. D. (2012). Big Data: Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*, 7(1), 1–5.

Stienmetz, J. L., & Fesenmaier, D. R. (2016). Validating Volunteered Geographic Information: Can We Reliably Trace Visitors´ Digital Footprints? In *Travel and Tourism Research Association: Advancing Tourism Research Globally* [online]. Retrieved from https://scholarworks.umass.edu/ttra/2016/Academic_Papers_Visual/24a.

Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, 70, 1–10. DOI: 10.1016/j.tourman.2018.07.010.

Tchetchik, A., Fleischer, A., & Shoval, N. (2009). Segmentation of Visitors to a Heritage Site Using High-resolution Time-space Data. *Journal of Travel Research*, 48(2), 216–229. DOI: 10.1177/0047287509332307.

The Guardian. (2012). *Hacker advertises details of 117 million LinkedIn users on darknet* [online]. Retrieved from https://www.theguardian.com/technology/2016/may/18/hacker-advertises-details-of-117-million-linkedin-users-on-darknet.

Thevenot, G. (2007). Blogging as a Social Media. *Tourism and Hospitality Research*, 7(3/4), 287–289. DOI: 10.1057/palgrave.thr.6050062.

Vaynerchuk, G. (2013). *Jab, Jab, Jab, Right Hook: How to tell your story in a noisy social world*. New York, USA: HarperCollins Publishers.

Versichele, M., de Groote, L., Bouuaert, M. C., Neutans, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, 44, 67–81. DOI: 10.1016/j.tourman.2014.02.009.

Von Mises, L. (1998). *Liberalismus* (Liberalism). Prague, Czech Republic: Ekopress.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46, 222–232. DOI: 10.1016/j.tourman.2014.07.003.

Weaver, A. (2008). When Tourists Become Data: Consumption, Surveillance, and Commerce. *Current Issues in Tourism*, 11(1), 1–23.

Wöber, K. (2007). Similarities in Information Search of City Break Travelers — A Web Usage Mining Exercise. In Sigala, M., Mich, L., Murphy, J. (Eds.), *Information and Communication Technologies in Tourism 2007* (pp. 77–86). Vienna: Springer. DOI: 10.1007/978-3-211-69566-1_8.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179–188. DOI: 10.1016/j.tourman.2009.02.016.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. DOI: 10.1016/j.tourman.2016.10.001.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634–639. DOI: 10.1016/j.chb.2010.04.014.

Yu, S. C., Choi, W. W., Shin, D. B., & Ahn, J. W. (2014). A Study on Concept and Services Framework of Geo-Spatial Big Data. *Journal of Korea Spatial Information Society*, 22(6), 13–21. DOI: 10.12672/ksis.2014.22.6.013.

Zach, F., & Gretzel, U. (2011). Tourist-Activated Networks: Implications for Dynamic Bundling and EN Route Recommendations. *Information Technology & Tourism*, 13(3), 229–238.

Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place? *Tourism Management*, 59, 267–280. DOI: 10.1016/j.tourman.2016.08.009.

Zoltan J., & McKercher, B. (2015). Analysing intra-destination movements and activity participation of tourists through destination card consumption. *Tourism Geographies*, 17(1), 19–35. DOI: 10.1080/14616688.2014.927523.