This is a peer-reviewed, accepted author manuscript of the following research article: Hartmann, K., & Lederer, M. (2021). The current state of Big Data research in tourism: results of a systematic literature analysis. *Zeitschrift für Tourismuswissenschaft*, *13*(2), 209-226. https://doi.org/10.1515/tw-2021-0015

Kim Hartmann*, Matthias Lederer

The Current State of Big Data Research in Tourism: Results of a Systematic Literature Analysis

*Corresponding author

Abstract / Outline

The use of large and diverse data in real time (called Big Data) affects many business processes and models. The tourism industry, characterized by manifold sub-sectors and players, provides a variety of starting points for Big Data usage. Examples are the optimization of transport offers using transaction data or a comprehensive analysis of destination trends based on social media posts. Big Data is a trending topic, however, the general discourse centres around potential ideas but fewer practical solutions. Based on a systematic literature analysis of initially 148 peer-reviewed journal articles, this article evaluates the current state of Big Data research within tourism. For this purpose, research articles centring around tourism-related Big Data were investigated according to the actual state of implementation of an IT solution, whether they truly grasp or represent Big Data in technological terms, and which added value they create for the tourism industry and research community. One key finding is that traditional data analysis is often wrongfully subsumed under the Big Data label. Further, the scientific literature predominantly discusses ideas or theoretical considerations, fewer tangible Big Data implementations, and fails to address and/or meet all requirements to be classified as Big Data. Only a minority of the presented solutions processes data in real time, whereas many rely on only one data source or structured data. Furthermore, most articles revolve around post-trip data analyses and are set to a destination context. In contrast, other tourism sectors as well as data interpretation and usage in pre-trip and on-trip phases are less represented. Lastly, this literature analysis provides an overview of true Big Data solutions already in operation and enables researchers to validly classify their own research activities in order to plan initiatives more specifically.

Keywords: tourism, big data, big data requirements, literature analysis

1 Introduction

The analysis of particularly large amounts of data with new methods (e.g. technical tools or algorithms) has been referred to as Big Data for quite some time (Chen et al., 2014). Especially in combination with other innovative concepts (e.g. process automation, self-service systems, individualisation of products and services), companies can benefit from a variety of application possibilities (Lin, 2014) to holistically increase efficiency and/or effectiveness of their business practices (Batistič & der Laken, 2019). Existing digital processes constitute one prevailing principle to facilitate the adoption of Big Data solutions in everyday business, as such processes create the required input of internal and external data (e.g. from customers or service partners) across a company's touchpoints and activities (Holmlund et al., 2020). Based on these specifications, the tourism industry is predestined to benefit from Big Data solutions, as it not only comprises manifold companies and establishments from all sectors related to attractions, facilities, and services for tourists but is further shaped by a high level of interaction as well as interdependence between individual industry players (Medlik, 2002; Pizam, 2009). Thus, a vacation or tourism experience often includes cooperation between a variety of services and touchpoints across tourism sectors, such as tour operators and travel agencies, transportation, destinations, accommodation, attractions, and activities (Cooper et al., 2008; Holloway et al., 2009).

It can be said that the actual vacation is not fully digitised - at least not yet by default (Happ & Ivancsó-Horváth, 2018). However, data is collected across numerous tourism-centred processes. This includes transaction processes (e.g. configuration, booking, payment) but also internal business processes of involved parties. Likewise, external data (e.g. social media reviews, semantic information from Google) proves to be increasingly important for the strategic management of tourism companies (Mariani et al., 2018). It is widely recognised, that the integration as well as systematic and timely analysis of vast amounts of data potentially provides key contributions to successful tourism practice as it supports (i) an increase of customer satisfaction (e.g. Moro et al., 2019, on superior recommendations, and Basaran et al., 2020, on shorter response times) and (ii) the development of new business models (e.g. Önder, 2017, and Stylos & Zwiegelaar, 2019, on early identification of travel trends and tapping into destinations as first movers).

Nevertheless, the implementation and use of Big Data is accompanied by various technical and organisational hurdles (Xiang, 2018; Iafrate, 2015). From a technological perspective, problems arise if certain data formats cannot be successfully integrated, known algorithms do not allow large-scale real-time analysis, or new user interfaces have yet to be developed (Stryk, 2015). Organisational challenges include management questions, for example for or against the implementation of new findings in existing solutions and the translation of these insights into specific ideas to design new offers, to adapt product or service ranges, or to smoothen internal and external processes (Xiang, 2018). For both problem areas, potential solutions can be expected from software manufacturers and from researchers. Since the underlying transaction and business processes are commonly realised via software, it is only logical that a large number of classic software manufacturers are setting out to use Big Data as a selling point (e.g. SAP, ORACLE). Likewise, numerous experimental, conceptual, and constructivist research outputs give guidance on how to make Big Data manageable and applicable in a tourism context. These range from rough ideas (e.g. Del Vecchio et al., 2018) to realised and evaluated prototypes (e.g. Fuchs et al., 2014).

However, numerous studies point out that not every research article or software solution titled with Big Data actually complies with a “real” Big Data approach (e.g. Lederer & Lederer, 2021). Rather, the term sustained a recognised hype to increase marketing potentials and is often utilised as an attention-seeking buzzword (Gandomi & Haider, 2015). This paper sets out to explore if tourism-related Big Data research truly grasps the nature and associated requirements of the Big Data phenomenon, or if the field suffers from an exploitation of the Big Data label for its own sake. In addition, an overview of the current stream of research on Big Data in tourism-related studies is provided. For this purpose, the following research questions are addressed:

> Do current tourism-related research contributions labelled with the term Big Data fulfil the underlying technical and additional requirements of Big Data? What are common shortcomings of non-complying contributions?
>
> Which Big Data solutions are already in operation in tourism contexts and provide a potential transferability to other tourism sectors?
>
> Is the field of tourism-related big data research fairly distributed across tourism sectors and trip phases? If not, which identified gaps remain to be covered in the future?
>
> How can researchers plan future contributions in a specific and meaningful way both for industry and academia?

Based on these research questions, this study aims to describe the current state of tourism-related Big Data research as an overview study. It has to be noted that there is a growing body of research on tourism-related Big Data solutions. However, past studies have mainly yielded important insights on deriving requirements for implementing companies (e.g. Vinod, 2016) and on the effect of Big Data on selected parameters such as performance (e.g. Stylos et al., 2021) or customer insights (e.g. Xu et al., 2020; Centobelli & Ndou, 2019). This contribution aims to complement the existing body of knowledge, as it is not limited to specific functions or areas of tourism (e.g. Petcu, 2011) and/or specific methods of Data Science (e.g. Alaei et al., 2019; Shapoval et al., 2018). Rather, it creates a comprehensive meta-view that complements the existing findings on applications (e.g. Donovan et al., 2017) as well as literature review studies such as Mariani et al. (2018) with a sound and critical assessment of the status quo.

First, the priorly mentioned requirements for Big Data are defined based on recognised definitions. Research contributions are then systematically reviewed to see whether and to which extent they include and/or fulfil the required set of attributes. Following the analysis, the identified degrees of fulfilment are described and discussed, resulting in a comprehensive overview of the state of the art in tourism-related Big Data research.

2 Method

The following analysis of tourism-related Big Data studies requires two methodical steps: (i) First, a definition of the characterising Big Data requirements for studies to be considered truly as Big Data research. (ii) Then, the systematic check of these requirements against a large number of scientific research articles. The quantitative evaluation enables an overview of the general scope and manifestations of big data research in tourism and is displayed in the "Results" section. The qualitative evaluation within this paper's "Discussion" section provides more detailed findings and an in-depth analysis of the field.

2.1 Big Data Requirements

The literature contains many documented attempts to describe Big Data and to isolate requirements for respective software solutions. To determine the underlying set of requirements for this paper, a five-step literature review was conducted following Mayring (2000). Moreover, the recognised approaches of Webster and Watson (2002) as well as Booth et al. (2016) were used to further generalise requirements: (i) The research question was translated into problem statements for the database search procedure. (ii) Academic publications were then identified in various databases to provide requirements. Even though textbooks tend to suffer from simplification, they were considered useful for the purpose of generalising easy-to-test requirements that are in-line with current streams of research. Therefore, also a total of 12 textbooks were explicitly considered. The search was conducted in German and English documents. (iii) After the search, publications were evaluated, analysed, and interpreted according to the priorly mentioned approaches. The included requirements generally follow the 3V concept (Volume, Velocity, Variety of Big Data) or extended versions (e.g. the widespread 5V model according to Gupta & Gupta, 2016). (iv) Lastly, and following Pohl (2010), a categorisation of information led to the textual specification of final Big Data requirements. Based on this, four key Big Data requirements were derived and described in general terms. This consolidated form is intended to serve as a compact and practicable assessment set and is thus recognised both in science (e.g., Bhat & Quadri, 2015) and teaching practice (e.g. Laudon & Laudon, 2017) for reducing the complexity of Big Data:

1. **Source**: An IT solution must potentially be able to feed from different data sources and to gain insights from a combination of those sources. This combination can encompass internal and external sources (e.g. own sales figures and Facebook trends) as well as horizontal/vertical integration of own databases (e.g. automatic merging of hotel ratings and customer feedback after the trip). This criterion is also fulfilled if not only micro-data, but rather meso- (e.g. contingency factors) or macro-data (e.g. open data streams) are processed.
2. **Speed**: The data analyses must be available and analysed in an appropriate time span for the scenario. A reference to waiting time because of memory structures or batch processing is not acceptable. The target rather is real-time analysis to allow immediate (re-)action.
3. **Value**: The discussed research must provide a significantly better solution for strategic or operational business problems with the help of Big Data. As described in the introduction, this can be to reach business goals based on data or to make proper use of the potentials of Big Data. In a nutshell, this requirement asks for an increase in added value, be it for a company, customer(s), or business partners.
4. **Processing**: Big Data does not rest on the use of quantitative standard procedures of descriptive analysis (e.g. summation, relative distribution). Rather, it explicitly requires the application of advanced methods (e.g. K-means clustering, association rule mining, K-nearest neighbour, support vector machine, decision tree, Bayesian network, artificial neural networks). As an alternative, predictive (e.g. support vector regression, time series analysis) or prescriptive analytics (e.g. genetic algorithms, constrained Bayesian network, expert Systems) can also be used.

2.2 Selection criteria and procedure

In the second step, the priorly described set of requirements was used for the structured analysis and detailed evaluation of current research contributions in the field of interest. To enable the analysis, research contributions were collected via the EBSCOHost database. This database was used following Lederer (2019) who describes that it serves well (alongside IEEE and SpringerLink) as a sound scientific reference for technical meta-studies related to business applications. The specific query ("Big Data" AND "Tourism") results directly from the research question. After setting these search terms, the authors examined whether the upcoming results also included synonyms or sub-sectors of tourism (e.g. travel, hospitality) to include all articles of relevance. The assumption was confirmed and therefore no adjustment of search terms was needed. In September 2020, the database contained 148 articles meeting the query as well as further quality criteria following recommendations by Kornmeier (2007), such as English language, peer-review, sound scientific methodology. While no restrictions were applied on the underlying research methodology or nature of a study's outcome (e.g. theories, models, prototypes), it was verified that all contributions clearly claim to be Big Data contributions. This quality-assuring step resulted in an exclusion of 75 inadequate articles, based on the publications' lack of keywords and/or content related to Big Data.

The following in-depth analysis focused on the actual implementation of a Big Data solution in practice, and the tourism sector in focus: (i) tour operators and travel agencies, (ii) transportation, (iii) destinations incl. attractions, and (iv) accommodation. Further, the examined articles were scanned for the time at which data was (or was planned to be) analysed, interpreted, and/or used to enable Big Data procedures. The underlying categorisation followed the popular consideration of pre-, on-, and post-trip periods within tourism (Jun et al., 2007; Kandampully et al., 2018). Lastly, the publications were compared to the pre-set Big Data requirements based on the content of the journal article. The methodology of the test followed the "Target Activity Grid" procedure based on recommendations of Best and Weth (2010). The output was an evaluation matrix in which each individual requirement (n=4) is assessed for each individual publication (n=73) resulting in 292 binary coding. The actual evaluation was carried out in two steps in the style of a Delphi study: First, an expert gave an assessment (requirement fulfilled or not). Then another expert independently checked this assessment. The few existing deviations were discussed in the team.

3 Results

As priorly outlined, the tourism-related Big Data research articles were matched against the provision of an actual IT solution in practice as well as the four predefined Big Data requirements (source, speed, value, and processing) to facilitate a first quantitative overview of the general stream of research in the examined field. The results of this process are outlined subsequently.

Table 1 presents the top ten most cited articles included in this study, which make up a total of

971 citations. Despite the fact that tourism science is shaped by specific tourism-centred journals, four of the top ten articles are to be found in non-tourism journals (e.g. Miah et al., 2017; Athanasopoulos et al., 2011).

~ INSERT TABLE 1 ~

Half (50,7%) of the examined articles include an implemented IT solution in practice, as demonstrated in Figure 1. In comparison, the share of implemented IT solutions lies slightly higher in the previously outlined top ten articles (60,0%). At large, most contributions fail to include, address, and meet all underlying Big Data requirements. Figure 1 presents the corresponding applicability of each requirement within the 73 articles in question, illustrating that value (60,3%) constitutes the prevalent criterion, followed by processing (42,5%), source (34,2%), and speed (23,3%). Further, a third (32,9%) of the articles fail to meet any of the four requirements. In contrast, few (12,3%) publications meet all four of the pre-set criteria, whereas the midfield is shaped by the fulfilment of three (19,2%), two (17,8%), or one (17,8%) requirement(s).

~ INSERT FIGURE 1 ~

It can be inferred from Figure 2 that an article's level of requirement fulfilment is largely influenced by the actual inclusion of a practical IT solution. As priorly outlined, half (49,3%) of tourism-related Big Data articles do not include the implementation of an IT solution in practice. Only a smaller number of no-solution articles meet the requirements of source (8,3%), speed (13,9%), value (33,3%), and/or processing (19,4%), whereas solution-inclusive publications demonstrate substantially higher levels of requirement fulfilment across all criteria: source (59,5%), speed (32,4%), value (86,5%), and/or processing (64,9%).

~ INSERT FIGURE 2 ~

With regards to the tourism sector in question, most tourism-related Big Data research (36 articles) centres around destinations followed by accommodation, tour operators & travel agencies, and transportation (in descending order, cf. Figure 3).

~ INSERT FIGURE 3 ~

One third (31,5%) of contributions involved data processing in a post-trip phase (e.g. after tourists left a destination or hotel), whereas the pre- (23,3%) and on-trip (13,7%) stage are far less represented within tourism-related Big Data research. The remaining contributions (31,5%) did not specify the time of data collection, analysis, and usage.

4 Discussion

The findings of this study provide evidence that the field of tourism-related Big Data research is characterised by a balance of theoretical and solution-based publications. Theory-centred contributions tend to revolve around potential scenarios for the use of Big Data and result in models or a comparison of existing approaches, for example in the context of automated sentiment analysis (e.g. Alaei et al., 2019; Kirilenko et al., 2018), predictions and forecasting (e.g. Cheng & Zhao, 2019; Tsai, 2014), web 2.0 (e.g. Donovan et al., 2017; Lee et al., 2019; Vinod, 2016), and social media (e.g. Edwards et al., 2017; Park et al., 2016; Vaid, 2018). Nevertheless, it is striking, that the proposed models and approaches to implement Big Data in the tourism industry rarely result in an actual IT solution to test their applicability in practice. In contrast, those articles including an actual IT solution perform substantially better across the monitored Big Data requirements as they are more likely to include, address, and meet four (e.g. Huang et al., 2017; Nilashi et al., 2015; Sinha et al., 2018), or three (e.g. Basaran et al., 2020; Fronzetti Colladon et al., 2019; Pantano & Dennis, 2019; Ren et al., 2019) Big Data requirements. The examined IT solutions commonly revolve around the potential of Big Data practice considering smart cities or smart destinations (e.g. Del Vecchio et al., 2018, Min et al., 2018; Zhao & Hu, 2019), tourist behaviour and movement at a destination (e.g. Liu et al., 2018; Önder et al., 2016; Scuderi & Dalle Nogare, 2018; Tingting, 2019), recommendations or reviews (e.g. Fronzetti Colladon et al., 2019; Nilashi et al., 2015; Ren et al., 2019), and predictions or forecasting (e.g. Cheng & Zhao, 2019; Fronzetti Colladon et al., 2019; Pan & Yang, 2017; Song et al., 2013; Tsai, 2014; Yang et al., 2014).

A closer examination of the discourse on Big Data requirements throughout the field points out that the creation of additional value is a predominant driver of Big Data research. However, it remains questionable how to implement successful Big Data procedures without complying to the other three crucial prerequisites as well (cf. Fig. 1). Especially, the predominant focus on data collection, analysis, and use in the post-trip phase prevents the exploitation of real-time data at other relevant points of time, for example to facilitate customer-oriented upselling or recommendation measures. Given that operational IT processing in tourism is characterised by phenomena as dynamic pricing and dynamic packaging during the (pre-trip) booking phase (Gibbs et al., 2018), it is surprising, that no vast attempts are made to establish a sound connection between rigorous research and relevance in operational tourism practice.

With a focus on the tourism sector in question, it is apparent that Big Data research in tourism is primarily shaped by destination-focused research. This circumstance is in-line with the nature of tourism research in general. Within the examined articles on Big Data in the context of destinations, one can subsume recurring key topics such as the analysis of visitor numbers and forecasts (e.g. Shapoval et al., 2018; Song et al., 2013), activities and attractions visited at the destination (e.g. Ren et al., 2019; Liu et al., 2018; Önder et al., 2016; Scuderi & Dalle Nogare, 2018; Tingting, 2019), as well as motivation to visit and destination choice (e.g. Lee et al., 2019; Wu et al., 2014). In contrast, those articles illustrating Big Data potentials for accommodation providers mostly highlight the applicability to forecast occupancy rates and demand (e.g. Pan & Yang, 2017; Yang et al., 2014), to increase firm performance (e.g. Yadegaridehkordi et al., 2020), and to evaluate potentially successful hotel locations (e.g. Mindur, 2019; Shereni & Chambwe 2020). In addition, tour operators and (online) travel agencies are represented within Big Data research. Respective articles address areas such as sustainable supply chains (e.g. Mandal, 2018), the sectors of online travel agencies or booking engines (e.g. Gibbs et al., 2018; Yao et al., 2019), and travel forums (e.g. Basaran et al., 2020; Fronzetti Colladon et al., 2019; Nilashi et al., 2015). The transportation sector constitutes the least represented sector in tourism-related Big Data research. This is surprising as its sub-sectors (e.g. aviation) deal with vast amounts of passenger, traffic and transactional data on a daily basis. The articles falling into this field mostly consider railway forecasting (e.g. Tsai, 2014; Hasuike & Ichimura, 2013), or self-driving activities and city traffic (e.g. Huang et al., 2017; Zhao & Hu, 2019). Again, discrepancies emerge between Big Data research and industry practice, as operational processes in the transportation sector rely heavily on real-time data. Further, the industry is shaped by an increased use of apps, for example in car sharing or aviation, which provide additional sources of rich and useful data.

From a technological perspective, Big Data research commonly involves detailed information on potential or exploited data source(s) (e.g. Li et al., 2018). The analysis reveals that Big Data research in tourism mainly relies on data from social media platforms (e.g. Del Vecchio et al., 2018), search engines (e.g. Önder, 2017), or geo-tags (e.g. Liu et al., 2018; Scuderi & Dalle Nogare, 2018). The publications are mostly bound to the use of one individual - and often external - data source instead of multiple sources to obtain relevant data. Further, research rests heavily on structured data that is little linked to other data sets. This fact leads to the absence of innovative approaches such as natural language processing (NLP) within the current state of research. Thus, even though a permanent synchronisation of multiple internal and external data sources across value chains and touchpoints can be considered as a worthwhile task from an economic perspective, the approach is scarcely suggested or adopted.

As priorly outlined, the field is characterised by many models and suggestions to be potentially applied in a tourism context. In contrast, only few publications involve actual IT solutions combined with high levels of fulfilment in all four Big Data requirements. Subsequently, the best performing contributions are emphasised, as they provide potential for the application in a variety of tourism sectors. The studies of Basaran et al. (2020), Fronzetti Colladon et al. (2019), and Nilashi et al. (2019) provide substantial insights in the successful use of Big Data solutions in the hospitality and aviation industry based on data from travel forums, such as HolidayCheck and Trip Advisor. In detail, Nilashi et al. (2019) applied a collaborative filtering approach to provide hotel recommendations based on preference learning, whereas Basaran et al. (2020) utilised an attribute-based approach to evaluate and forecast hotel performance. Fronzetti Colladon et al. (2019) apply user-generated content (UGC) to forecast airport arrivals in major European cities and recommend the solution to be integrated in traditional forecasting models. The potential of using multiple data sources is demonstrated in Pan and Yang's (2017) publication based on combined data from search engine queries, website traffic, and weekly weather information. The resulting Big Data solution not only forecasts hotel occupancy rates, but also facilitates benchmarking and the optimisation of hotel operations (Pan & Yang, 2017). Lastly, Min et al. (2018) outline how Big Data solutions can solve imbalances between tourist distribution and attraction capacity at a destination level by using Big Data learnings from city traffic contexts. A consideration of the outlined studies and their transferability to other tourism sectors is recommended both for tourism researchers and practitioners.

5 Summary

To subsume, the limitations of this paper need to be acknowledged. As the findings rely on one academic database, this paper is not representative for the whole field of tourism-related Big Data research. Nevertheless, the authors agree that this literature analysis represent a variety of both technical and tourism-related perspectives on Big Data, and is in line with the scope of comparable publications (e.g. Mariani et al., 2018). Therefore, it provides rich insights on the current stream of research in this area. Another bias results from the pronounced focus on the destination sector, resulting in an underrepresentation of other sectors of the tourism industry. This distribution might result from the general preponderance of destination-related publications within tourism research, from the widespread conception of tourism as a sub-category of geography, and from the availability of research funds in this sector.

This literature analysis contributes to the field in numerous points. First, it seems that tourism research often tends to use the Big Data label without adopting and/or meeting the substantial Big Data prerequisites of source, speed, value, and processing. Second, even though the literature is characterised by a balance of theoretical and solution-oriented publications, the manifestations rank between niche-oriented (e.g. eco-friendly hotels) and all-encompassing reflections (e.g. tourism in general). Lastly, the analysed research spectrum mostly fails to combine potentials of Big Data with current developments in tourism practice (e.g. dynamic pricing, integration of data sources and system interfaces). Future research will have to continue to explore the implementation of actual Big Data solutions in practice and to cater to the variety of tourism sectors. Further, this study exposed the need to comply with key Big Data requirements and, especially, to include and combine real-time transactional data of internal and external data sources. In agreement with Mariani et al. (2018), this shortcoming is to be dissolved by an encouragement of multi-disciplinary research teams comprising both tourism and IT expertise.

In practice, Big Data reveals considerable potential to add substantial value across all participants along the tourism value chain. This potential should be mirrored by future streams of research, ensuring rigour, relevance, and applicability.

References

Athanasopoulos, G., Hyndman, R.J., Song, H., & Wu, D.C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822-844.

Alaei, A.R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.

Basaran, M.A., Dogan, S., & Kantarci, K. (2020). On modeling of responses generated by travel 2.0 implementation: Fuzzy rule-based systems. *International Journal of Contemporary Hospitality Management*, 32(4), 1503-1522.

Batistič, S., & der Laken, P. (2019). History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. *British Journal of Management*, 30(2), 229-251.

Best, W., & Weth, M. (2010). *Process Excellence*. Cham: Springer.

Bhat, W., & Quadri, A. (2015). Big Data promises value: is hardware technology taken onboard?. *Industrial Management & Data Systems*, 115(9), 1577-1595.

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review*. Los Angeles: SAGE.

Centobelli, P., & Ndou, V. (2019). Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism*, 22(15), 1862-1882.

Chen, M., Mao, S., Zhang, Y., & Leung, V.C.M. (2014). *Big data: Related technologies, challenges and future prospects*. Cham: Springer.

Cheng, X., & Zhao, C. (2019). Prediction of tourist consumption based on bayesian network and big data. *Ingénierie Des Systèmes D'Information*, 24(5), 491-497.

Cooper, C., Fletcher, J., Fyall, A., Gilbert, D., & Wanhill, S. (2008). *Tourism: Principles and practices* (4th ed.). Essex: Pearson Education.

Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management*, 54(5), 847-860.

Donovan, C., Flaherty, E.T., & Quinn Healy, E. (2017). Using big data from Wikipedia page views for official tourism statistics. *Statistical Journal of the IAOS*, 33(4), 997-1003.

Edwards, D., Cheng, M., Wong, I.A., Zhang, J., & Wu, Q. (2017). Ambassadors of knowledge sharing: Co-produced travel information through tourist-local social media exchange. *International Journal of Contemporary Hospitality Management*, 29(2), 690-708.

Fronzetti Colladon, A., Guardabascio, B., & Innarella, R. (2019). Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems*, 123, 113075.

Fuchs, M., Höpken, W., & Lexhagena, M. (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management*, 3(4), 198-209.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.

Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management*, 30(1), 2-20.

Gupta, U.G., & Gupta, A. (2016). Vision: A missing key dimension in the 5V big data framework. *Journal of International Business Research and Marketing*, 1(3), 50-56.

Happ, E., & Ivancsó-Horváth, Z. (2018). Digital tourism is the challenge of future – a new approach to tourism. *Knowledge Horizons - Economics*, 10(2), 9-16.

Hasuike, T., & Ichimura, T. (2013). Web intelligence for tourism using railway data by a simplified fuzzy reasoning method. *Journal of Intelligent & Fuzzy Systems*, 24(2), 251-259.

Holloway, J.C., Humphreys, C., & Davidson, R. (2009). *The business of tourism* (8th ed.). Essex: Pearson Education.

Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Ravald, A., Sarantopoulos, P., Villarroel Ordenes, F., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research*, 116, 356-365.

Huang, Z., Cao, F., Jin, C., Yu, Z., & Huang, R. (2017). Carbon emission flow from self-driving tours and its spatial relationship with scenic spots – A traffic-related big data method. *Journal of Cleaner Production*, 142, 946-955.

Iafrate, F. (2015). *From big data to smart data*. London: Wiley.

Jun, S.H., Vogt, C.A., & MacKay, K.J. (2007). Relationships between travel information search and travel product purchase in pretrip contexts. *Journal of Travel Research*, 45(February), 266-274.

Kandampully, J., Zhang, T., & Jaakkola, E. (2018). Customer experience management in hospitality: A literature synthesis, new understanding, and research agenda. *International Journal of Contemporary Hospitality Management*, 30(1), 21-56.

Kirilenko, A.P., Stepchenkova, S.O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 57(8), 1012-1025.

Kornmeier, M. (2007). *Wissenschaftstheorie und wissenschaftliches Arbeiten*. Physica, Mannheim.

Laudon, C., & Laudon, J.P. (2017). *Management Information Systems: Managing the Digital Firm*. London: Pearson.

Lederer, M. (2019). What's going to happen to Business Process Management? Current Status and Future of a Discipline. In Proceedings of the S-BPM ONE 2019. Sevilla: CEUR-WS.

Lederer, M., & Lederer J. (2021): Are they Ready for the Big Thing? Big Data Applications Requirements for Process Management & Evaluation of Current Software Solutions. In: J. P. Chigwada & G. Tsvuura (Hrsg.): *Handbook of Research on Information and Records Management in the Fourth Industrial Revolution*. Hershey: Information Science Reference.

Lee, H., Chung, N., & Nam, Y. (2019). Do online information sources really make tourists visit more diverse places? Based on the social networking analysis. *Information Processing & Management*, 56(4), 1376-1390.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323.

Lin, N. (2014). *Applied Business Analytics: Integrating Business Process, Big Data, and Advanced Analytics*. Jew Jersey: FT Press.

Liu, Q., Wang, Z., & Ye, X. (2018). Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Transactions in GIS*, 22(6), 1372-1389.

Mandal, S. (2018). Exploring the influence of big data analytics management capabilities on sustainable tourism supply chain performance: The moderating role of technology orientation. *Journal of Travel & Tourism Marketing*, 35(8), 1104-1118.

Mariani, M., Baggio, R., Fuchs, M. & Höpken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 30(12), 3514-3554.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung*, 1(2).

Medlik, S. (2002). *Dictionary of travel, tourism, and hospitality* (3rd ed.). Oxford: Elsevier.

Miah, S.J., Vu, H.Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771-785.

Min, W., Yu, L., Yu, L., & He, S. (2018). People logistics in smart cities. *Communications of the ACM*, 61(11), 54-59.

Mindur, M. (2019). The determination of potential locations for hotel and service facilities in relation to the transport system - the logistic approach. *Logistics & Transport*, 43(3), 27-35.

Moro, S., Esmerado, J., Ramos, P., & Alturas, B. (2019). Evaluating a guest satisfaction model through data mining. *International Journal of Contemporary Hospitality Management*, 32(4), 1523-1538.

Nilashi, M., Ahani, A., Esfahani, M.D., Yadegaridehkordi, E., Samad, S., Ibrahim, O., Sharef, N.M., & Akbari, E. (2019). Preference learning for eco-friendly hotels recommendation: A multi-criteria collaborative filtering approach. *Journal of Cleaner Production*, 215, 767-783.

Nilashi, M., Bin Ibrahim, O., Ithnin, N., & Sarmin, N.H. (2015). A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS. *Electronic Commerce Research and Applications*, 14(6), 542-562.

Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints. *Journal of Travel Research*, 55(5), 566-573.

Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *The International Journal of Tourism Research*, 19(6), 648-660.

Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56(7), 957-970.

Pantano, E., & Dennis, C. (2019). Store buildings as tourist attractions: Mining retail meaning of store building pictures through a machine learning approach. *Journal of Retailing and Consumer Services*, 51, 304-310.

Park, S.B., Ok, C.M., & Chae, B.K. (2016). Using Twitter data for cruise tourism marketing and research. *Journal of Travel & Tourism Marketing*, 33(6), 885-898.

Petcu, N. (2011). Statistical tests applied in tourism. *Economic Sciences*, 4(1), 135-140.

Pizam, A. (2009). *What is the hospitality industry and how does it differ from the tourism and travel industries?* International Journal of Hospitality Management, 28, 183-184.

Pohl, K. (2010). *Requirements Engineering: Fundamentals, Principles, and Techniques*. Cham: Springer.

Ren, Q., Xu, F., & Ji, X. (2019). Use of the pathfinder network scaling to measure online customer reviews: A theme park study. *Strategic Change*, 28(5), 333-344.

Scuderi, R., & Dalle Nogare, C. (2018). Mapping tourist consumption behaviour from destination card data: What do sequences of activities reveal?. *The International Journal of Tourism Research*, 20(5), 554-565.

Shapoval, V., Wang, M.C., Hara, T., & Shioya, H. (2018). Data mining in tourism data analysis: Inbound visitors to Japan. *Journal of Travel Research*, 57(3), 310-323.

Shereni, N.C., & Chambwe, M. (2020). Hospitality big data analytics in developing countries. *Journal of Quality Assurance in Hospitality & Tourism*, 21(3), 361-369.

Sinha, S., Bhatnagar, V., & Bansal, A (2018). Multi-label naïve bayes classifier for identification of top destination and issues to accost by tourism sector. *Journal of Global Information Management*, 26(3), 37-54.

Song, H., Gao, B.Z., & Lin, V.S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29(2), 295-310.

Stryk, B. (2015). *How do organizations prepare and clean big data to achieve better data governance? A Delphi Study*. Dissertation, Capella University.

Stylos, N., & Zwiegelaar, J. (2019). Big data as a game changer: How does it shape business intelligence within a tourism and hospitality industry context? In: Marianna Sigala, Roya Rahimi, Mike Thelwall (Eds.): *Big Data and Innovation in Tourism, Travel, and Hospitality* (p.163-181). Cham: Springer.

Tingting, S. (2019). Spatial data mining and big data analysis of tourist travel behavior. *Ingénierie Des Systèmes D'Information*, 24(2), 167-173.

Tsai, T. (2014). A self-learning advanced booking model for railway arrival forecasting. *Transportation Research. Part C, Emerging Technologies*, 39, 80-93.

Vaid, J. (2018). Role of big data analytics in social media marketing of MICE tourism. *Global Journal of Enterprise Information System*, 10(1), 55-62.

Vinod, B. (2016). Big data in the travel marketplace. *Journal of Revenue and Pricing Management*, 15(5), 352-359.

Webster, J., Watson, R.T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26, 13-23.

Wu, H.-Q., Liu, Y.-X., Zhuang, D.-L., & Zhu, W.-D. (2014). A DEA-based measurement of effectiveness of provincial image advertisement for local tourism destination: Evidence from China. *International Journal of Management & Decision Making*, 13(2), 192-207.

Xiang, Z. (2018). From digitization to the age of acceleration: On information technology and tourism. *Tourism Management Perspectives*, 25, 147-150.

Xu, F., Nash, N., & Whitmarsh, L. (2020). Big data or small data? A methodological review of sustainable tourism. *Journal of Sustainable Tourism*, 28(2), 147-166.Yadegaridehkordi, E., Nilashi, M., Shuib, L., Hairul Nizam Bin Md Nasir, M., Asadi, S., Samad, S., & Fatimah Awang, N. (2020). The impact of big data on firm performance in hotel industry. *Electronic Commerce Research and Applications*, 40, 100921.

Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433-447.

Yao, B., Qiu, R.T.R., Fan, D.X.F., Liu, A., & Buhalis, D. (2019). Standing out from the crowd – an exploration of signal attributes of Airbnb listings. *International Journal of Contemporary Hospitality Management*, 31(12), 4520-4542.

Zhao, P., & Hu, H. (2019). Geographical patterns of traffic congestion in growing megacities: Big data analytics from Beijing. *Cities*, 92, 164-174.