

Business intelligence and big data in hospitality and tourism: a systematic literature review

Marcello Mariani

Henley Business School, University of Reading, Reading, UK

Rodolfo Baggio

Dondena Center for Research on Social Dynamics and Public Policy, Bocconi University, Milan, Italy and National Research Tomsk Polytechnic University, Tomsk, Russia

Matthias Fuchs

Mid-Sweden University, Östersund, Sweden, and

Wolfram Höepken

University of Applied Sciences Ravensburg-Weingarten, Weingarten, Germany

Abstract

Purpose – This paper aims to examine the extent to which Business Intelligence and Big Data feature within academic research in hospitality and tourism published until 2016, by identifying research gaps and future developments and designing an agenda for future research.

Design/methodology/approach – The study consists of a systematic quantitative literature review of academic articles indexed on the Scopus and Web of Science databases. The articles were reviewed based on the following features: research topic; conceptual and theoretical characterization; sources of data; type of data and size; data collection methods; data analysis techniques; and data reporting and visualization.

Findings – Findings indicate an increase in hospitality and tourism management literature applying analytical techniques to large quantities of data. However, this research field is fairly fragmented in scope and limited in methodologies and displays several gaps. A conceptual framework that helps to identify critical business problems and links the domains of business intelligence and big data to tourism and hospitality management and development is missing. Moreover, epistemological dilemmas and consequences for theory development of big data-driven knowledge are still a terra incognita. Last, despite calls for more integration of management and data science, cross-disciplinary collaborations with computer and data scientists are rather episodic and related to specific types of work and research.

Research limitations/implications – This work is based on academic articles published before 2017; hence, scientific outputs published after the moment of writing have not been included. A rich research agenda is designed.

Originality/value – This study contributes to explore in depth and systematically to what extent hospitality and tourism scholars are aware of and working intently on business intelligence and big data. To the best of the authors' knowledge, it is the first systematic literature review within hospitality and tourism research dealing with business intelligence and big data.

Keywords Big data, Tourism, Systematic literature review, Hospitality, Business intelligence

Paper type Literature review



1. Introduction

In recent years, the notion of “Big Data” (BD) has become increasingly popular in both academic and non-academic media. The concept has generated a real buzz, especially on the

internet and social media. At the moment of writing the paper at hand (February 2017), a Google search using the circumlocution “Big Data” obtained more than 290 million resulting items. The circumlocution identifies the enormous volume of both unstructured and structured data generated by technology developments and the exponentially increasing adoption of devices allowing for automation and connection to the internet. The use of networked devices, such as tablets and smartphones, has brought an explosion of data (Verhoef *et al.*, 2016), often in connection with user-generated content stemming from online social networks (Leung *et al.*, 2013). Given their volume and characteristics, BD are difficult to process by deploying traditional statistical methods and software techniques (Chen *et al.*, 2014). Nevertheless, BD is becoming rapidly popular as an emerging new field of inquiry also in social sciences, where it has been identified as an irreplaceable factor to enhance economic growth and prosperity and to solve societal problems (Mayer-Schönberger and Cukier, 2013) as well as a major driver for the creation of value for firms and customers (Verhoef *et al.*, 2016).

Methodologically, BD-based approaches allow researchers to overcome the difficulties of working with representative samples, as BD virtually allows working with the entire population under scrutiny (Gerard *et al.*, 2016). Supposedly, it enables to answer any question related to people’s opinions, views, ideas and behaviours. At the same time, it seems to be a powerful tool to address novel research questions, to develop innovative research designs useful in the advancement of knowledge, ultimately generating both policy and managerial decision support (Gerard *et al.*, 2016). On the one hand, there is consensus among business leaders and scholars, that BD represents a necessary condition to investigate today’s complex business and social phenomena through the possibility of combining and recombining extremely different sources of information (Bedeley and Nemati, 2014). Similarly, firms leveraging BD can enhance their competitive advantage in a world where markets are global and huge amounts of information about consumers are available on the internet (Verhoef *et al.*, 2016). On the other hand, BD brings along a significant number of new challenges, risks and dilemmas, which have been explored and dealt with in a number of works (Boyd and Crawford, 2012; Fan *et al.*, 2014; McFarland and McFarland, 2015; Ekbja *et al.*, 2015; Gerard *et al.*, 2016). Interestingly, scholars have addressed not only the challenges of data sharing and privacy, or new epistemological dilemmas but also the challenges related to data extraction, collection, storage, processing, analysis and visualization and reporting (Gerard *et al.*, 2016). These processes all require specific resources and specialized skills, which will increasingly be juxtaposed to more established research methods (Kitchin and Lauriault, 2015). Despite the aforementioned challenges, there is broad consensus within both academic and business circles, that BD could make a difference as it captures real-time behaviours and opinions on virtually any aspect of human life (Chang *et al.*, 2014; Power, 2014). For instance, in 2012, an IBM-based survey conducted on chief marketing officers revealed that BD is considered a major business challenge (IBM, 2012). As such, BD have also started to be a significant source for business intelligence (BI) activities aimed at creating, delivering and capturing customer value (Verhoef *et al.*, 2016).

BI and especially BI analytics have a longer tradition, but they constantly incorporate developments within data science in general and BD in particular, with the aim to enhance the return on investment and drive marketing decision-making (<https://cmosurvey.org/results/>). Consequently, both subjects are highly complementary. While the first known usage of the term BI dates back to 1865 (when the Banker Sir Henry Furness had the ability to profit from his remarkable understanding of political and market issues and instabilities before his competitors (Devens and Miller, 2013), the modern phase of BI dates back to the 1990s. This first turning point in the development of BI was characterized by tools

specialized in extracting, transforming and loading data into a central data store. These tools could also organize, visualize and descriptively analyse data, such as doing online analytical processing (OLAP). However, these tools were developed with anyone but experts in mind; thus, most users were not capable of carrying out BI tasks on their own. The exponential growth of the internet in the twenty-first century advanced this development and fully addressed the issues of complexity and speed. In particular, it addressed those issues through real-time solutions, cloud-based self-service options to improve data visualization as well as through advanced analytics and mobile-empowered BI platforms that integrated the end-user even more. BI is no longer an added utility, rather, it became a requirement for businesses looking to stay competitive, and even to remain afloat, in an entirely new, data-driven environment (Liebowitz, 2013).

The term BD appeared for the first time in Bryson *et al.*'s (1999) seminal paper published by the Communications of the ACM. Especially Web 2.0 applications and the rise of mobile devices further increased data volumes. Thus, BD is not anymore an isolated phenomenon, but one that is part of a long evolution of capturing and using data, both for societal, scientific and business purposes. BD analytics is leading to better and more informed decision-making for individuals and organizations and, especially, creates value for stakeholders and customers (Verhoef *et al.*, 2016).

Over the past decades, the fields of tourism, travel, hospitality and leisure have widely recognized the need for a *customer-centric* approach that primarily values tourists' needs, wants, preferences and requirements as major determinants in travel decisions to enhance both consumer satisfaction and the quality and memorability of the tourist experience (Correia *et al.*, 2013; Prayag *et al.*, 2013). Only very recently an increasing amount of work related to the fields of BI and BD has grown to enrich these two lines of research. To the best of our knowledge, no review study has previously explored in depth to what extent hospitality and tourism scholars are aware of and working intently on BI and BD. This is clear when observing that research on the role of BD and BI is still highly fragmented and relegated to isolated research questions. To address this research gap, we conduct a systematic quantitative literature review on the concepts of BI and BD and their application (and related techniques) in the fields of tourism and hospitality. To achieve this aim, our review is structured as follows: in Section 2, we provide a literature review of the wide field of BD and BI. In Section 3, we illustrate the research methodology adopted to conduct the systematic literature review. Section 4 illustrates the major findings of the review and describes the articles by identifying research topics, conceptual and theoretical approaches, research designs, methods for data collection, analysis and reporting/visualization, data features such as sources of data, type of data and size. In Section 5, we identify theoretical and methodological knowledge gaps and development needs in travel, hospitality and tourism, as well as promising research areas for BD and BI in hospitality and tourism. Finally, in Section 6, we draw major conclusions and discuss the limitations of our review.

2. Business intelligence and big data

2.1 Business intelligence

BI comprises all the activities, applications and technologies needed for the collection, analysis and visualization of business data to support both operative and strategic decision-making (Dedić and Stanier, 2016; Kimball and Ross, 2016). Currently, BI is used as an umbrella term to cover the domains of data warehousing (DW), data mining and reporting as well as OLAP (Williams, 2016). Already in 1958, IBM researcher Hans Peter Luhn used the term BI to highlight "*the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal*" (Luhn, 1958, p. 314). BI largely evolved

from the “decision support systems” (DSS) domain that emerged in the 1960s and developed throughout the mid-1980s. A DSS is defined as a computer-based information system that supports business or organizational decision-making activities that typically result in ranking, sorting or choosing from among alternatives (Burstein and Holsapple, 2008; Sauter, 2011). Classic DSS applications usually comprise computer-aided models, DW, OLAP and executive information systems (Kimball *et al.*, 2008). Only in 1989, later Gartner analyst Howard J. Dresner suggested that “*business intelligence*” is an umbrella term to describe “*concepts and methods to improve business decision making by using fact-based support systems.*” (Power, 2007, p. 128).

BI, as it is currently understood, offers historical, current and predictive views of business processes (Kimball *et al.*, 2008). Typical functions embrace reporting, OLAP, analytics, data mining, business performance management, benchmarking, text mining and prescriptive analytics (Rud, 2009; Williams, 2016). In particular, BI technologies show the capacity to handle large amounts of structured as well as unstructured data to help identify, develop or create new strategic business opportunities. Accordingly, enterprises apply BI to support a wide range of operational business decisions, such as product positioning or pricing. Moreover, BI provides strategic insights into new markets, supports the assessment of customer demand and the suitability of products and services developed for different market segments, or the impact of marketing and advertising strategies (Chugh and Grandhi, 2013). According to Kimball *et al.* (2008), BI is the most effective when it combines *external* data derived from the customer markets in which a company operates with *internal* data stemming from company sources, such as financial or booking data. Thus, when matched, external and internal data provide the most complete picture, creating “*intelligence*” that cannot be derived from any singular set of data (Coker, 2014).

However, businesses should assess three critical areas before implementing a BI project (Kimball *et al.*, 2008):

- the level of commitment and sponsorship of the BI project from senior management;
- the level of business needs for creating a BI implementation; and
- the amount and quality of business data available.

The latter requirement is an especially necessary condition, as without proper data (or with too low data quality) any implementation of a BI application will fail. Thus, *data profiling*, which aims at identifying the “*content, structure and consistency of data*” (Kimball *et al.*, 2008, p. 17), should happen as early as possible in the BI cycle. A synthetic, though, complete representation of BI and its articulation is provided in Table I, illustrating the major sub-components and fields related to BI.

2.2 Big data

BD has been portrayed as a contemporary hype (d’Amore *et al.*, 2015). A few scholars have described it in view of its popularization as a kind of panacea, able to provide a wealth of useful insights into many aspects of the life of individuals, organizations and markets (Mayer-Schönberger and Cukier, 2013; McAfee *et al.*, 2012). Other, more critical, scholars have depicted it as a cultural, technological and scholarly phenomenon that rests on the interplay of technology, analysis and even mythology (Boyd and Crawford, 2012; Ekbja *et al.*, 2015).

BD would not exist without technological development. Over the past three decades, the number of devices allowing for automation and/or connection to the internet has increased exponentially. These devices have brought about a proliferation of data, often in connection

Table I.
Major components
and functional
subdomains within
the BI umbrella

Major concept	Acronym	Short explanation	References
Decision support system	DSS	Computer-based information system that supports decision making resulting in ranking, sorting, or choosing from among alternatives	Burstein and Holsapple (2008), Sauter (2011)
Data warehousing	DW	Central data repository system of integrated data from one or multiple sources that stores current and historical data in one single place and format	Kimball <i>et al.</i> (2008)
Online analytical processing	OLAP	Provides multi-dimensional analytical queries encompassing DW and reporting. Supports the operations of consolidation, drill-down, slicing and dicing	Kimball <i>et al.</i> (2008)
Data mining	DM	Discovers correlations and patterns in (usually large) data sets involving methods of machine learning, statistics and mathematical modeling	Larose (2005), Rud (2009)
Business intelligence	BI	Umbrella term comprising the domains of DW, OLAP and DM	Kimball and Ross (2016)
Descriptive analytics	–	Uses data aggregation (e.g. sums, averages, percentages, changes, etc.) and data mining to provide insight into the past to answer: “What has happened?”	Williams (2016)
Predictive analytics	–	Uses statistical and DM models to forecast the future and answer: “What could happen?”	Dedić and Stanier (2016)
Prescriptive analytics	–	Uses machine learning and computational modelling to advice on optimal outcomes and answers: “What should we do?”	Dedić and Stanier (2016), Williams (2016)
Big Data	BD	Data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Includes challenges, as data extraction, storage, analytics, visualization, querying, updating and information privacy	Erl <i>et al.</i> (2015)

with user-generated content stemming from online social networks (Leung *et al.*, 2013; Mariani *et al.*, 2014), mostly accessible through a number of mechanisms and tools, such as application programming interfaces (APIs) (Russell, 2013; Mariani *et al.*, 2016). This explosion in data availability has attracted the attention of computer and data scientists, whose efforts, beyond traditional methods of data management and warehousing, have been directed to novel techniques for BD analysis (Franks, 2012). Most techniques can be classified as “Machine Learning” (ML), a term commonly deployed to define the methods and algorithms used for mining data with the aim of extracting patterns, correlations and knowledge from apparently unstructured data sources (Witten *et al.*, 2016). ML is undeniably gaining momentum as a set of processing techniques. Without going into details, it consists of learning algorithms discovering general rules or patterns in large data sets, filtering based on several variables, clustering large collection of objects into a small number of classes, etc. (Mitchell, 1997). Whether supervised (i.e. outcomes come from training the algorithms with pre-labelled data) or unsupervised (i.e. algorithms derive outcomes from the data themselves), these techniques aim at giving a computer the ability to

perform a task by using generalized approaches without being explicitly programmed for single tasks (for a more complete description: [Witten et al., 2016](#)).

It has been pointed out that BD displays a few main characteristics synthesized in the so-called “3Vs” ([Chen et al., 2014](#); [Laney, 2001](#)). Accordingly, BD is characterized in terms of “volume” (i.e. it cannot be stored in an ordinary PC, as it typically exceeds billions of gigabytes), “variety” (i.e. data comes in a wide range of forms and shapes, such as texts, sounds, pictures and videos and are often spatially and temporally referenced), and “velocity” (i.e. the speed at which data is created and modified is relevant). While the “v” of volume has caught most of the attention of scholars so far, the remaining two aspects, variety and velocity, also make BD particularly interesting to address a number of practical issues and research questions. More recently, [Baggio \(2016\)](#) has detected further features related to BD. More concretely, four additional “Vs” are recognized. “Value” is relevant as BD help to create value for individuals and organizations and is becoming an object of economic transactions themselves. Second, “variability” plays a major role as it comes in the form of unstructured records whose meaning varies across contexts and times. Third, “veracity” is related to the reliability, validity and completeness dimension of the data. Fourth, visualization, namely, the need to present the complexity of patterns in graphically understandable ways, is another important characteristic.

There are other features pertaining to BD. Often, whole populations, instead of samples, are gathered and empirically explored. This might lead to a reconsideration of statistical tools and inferential methods ([Fan et al., 2014](#)). These other features of BD include the following: the relational nature of variables that might be common across different sources; the flexibility needed when analysing the collections that might bring to scaling or easily extending variables and cases ([Mayer-Schönberger and Cukier, 2013](#)); the high probability to find a significant (but often spurious) correlation between any two series of data ([Granville, 2013](#)); and, finally, the high level of granularity, allowing to focus on specific features.

As far as technology is concerned, BD is often operated (when sizes are really big) through the Hadoop framework ([White, 2015](#)), which was generated in the form of an open source project of the Apache Software foundation, for both storage purposes and the distributed processing of data sets on clusters of commodity hardware. Distributed computing allows a virtually unlimited number of computers to process a significant amount of data simultaneously (in the order of petabytes, i.e. 10^{15} bytes). When it comes to the extraction of digital records, especially from third-party platforms, they can be retrieved quite easily through APIs that are freely accessible, for instance, for most social media platforms (e.g. Facebook, Twitter). They can be later processed through artificial intelligence (AI) methods, such as machine learning (ML) described above.

Leaving the technical aspects of BD (i.e. data retrieval, processing, analysis and visualization, respectively) aside, and going back to the characteristics of BD, their “value” is particularly relevant for business applications. BD can lead to better and more informed decision making for individuals and organizations and, thus, create value for stakeholders and customers ([Verhoef et al., 2016](#)). In other words, BD can be used for BI purposes. More specifically, BD can empower BI, but it needs to be elaborated in a proper way. It remains that the risk of discovering deceptive outcomes and effects is quite high, especially if the quality of data and data pre-processing is low ([Pyle, 1999](#); [Lazer et al., 2014](#)).

2.3 The connection between business intelligence and big data in hospitality and tourism

Any tourism company (be it a hotel or an airline) needs to leverage its managerial and marketing strategies, tactics and tools to achieve and maintain sustained competitive

advantage. This is more critical in the current highly dynamic economic environment where competition is fierce and consumers are demanding and experienced (Mariani and Baggio, 2012). Increasingly, it is evident that it is extremely difficult, even for well-established companies, to cultivate and sustain a competitive advantage for a long period. We are going through an age of “temporary advantage” and “hyper-competition” (D’Aveni, 1994; D’Aveni *et al.*, 2010), where organizations need constant innovation to gain a temporary benefit and move ahead of the competition for a continued series of time periods (Mariani *et al.*, 2016).

In this context, BD can make a difference for the BI of tourism companies, help them make better strategic and tactical decisions and create value (Verhoef *et al.*, 2016). This is the reason why BD is increasingly a crucial component of the wider BI umbrella (Section 2.1 and Table I). However, research on the role of BD for BI in the hospitality and tourism literature is still scant (d’Amore *et al.*, 2015; Baggio, 2016) and highly fragmented. Single research activities often take place in a rather isolated manner and tackle a very specific aspect or research question without looking at the whole picture and embedding new work into the overall scholarly and practical context. Such research practices, however, are common during the emergence of new research areas or phenomena (Knudsen, 2003). Therefore, in the current phase of development of hospitality and tourism research leveraging BD and BI, it is important and even overdue, to provide a clear overview of the different facets and issues of the wide research domain of BI and identify, discuss and integrate existing research activities leveraging BD into the overall context of the focal research domain. This is important in particular for two reasons: first, to stimulate and also to systematize further research activities; and second, to provide informational bases and overview on current application areas and utilization potentials for companies and stakeholders in the tourism domain. To achieve the aforementioned goals and bridge this gap, we analyse state-of-the-art work done in BI and BD in the domain of tourism and hospitality through a systematic literature review. In the next section, we explain the methodology adopted.

3. Methodology

To assess the extent to which BI and BD feature within the hospitality and tourism literature, we carried out a systematic literature review of academic articles indexed on the Scopus and Web of Science databases. The method of systematic literature review has been largely adopted in the wider social sciences (Tranfeld *et al.*, 2003), including in the hospitality and tourism domain (Gomezelj, 2016; Ip *et al.*, 2011; Law *et al.*, 2016). Thus, we embraced this approach to identify the relevant scientific work.

Subsequently, we have manually analysed and clustered refereed scientific articles into two major groups (i.e. BI-related and BD-related), and explored them in detail based on the following features:

- research topic;
- conceptual and theoretical characterization;
- sources of data;
- type of data and size;
- data collection methods;
- data analysis techniques; and
- data reporting and visualization.

The identification of the aforementioned features and clusters is done to help us identify theoretical and methodological knowledge gaps, development needs and promising research areas for both BI and BD in the hospitality and tourism domain.

3.1 Data

As mentioned, the present study retrieved data from two large databases: Scopus and Web of Science. The main reason for having selected the two aforementioned databases is that they are considered to be the most comprehensive sources of scholarly articles and academic work in the social sciences (Vieira and Gomes, 2009). More precisely, Scopus covers more than 22,000 titles from over 5,000 international publishers; therefore, it is considered one of the most comprehensive repository of the world's research output across a wide range of academic disciplines. On a similar scale, Web of Science (WOS) provides access to seven databases that reference cross-disciplinary research covering over 28,000 journals. To these sources, we juxtapose the digital library of the International Federation for Information Technologies, and Travel & Tourism (www.ifitt.org/resources/digital-library/) that indexes 992 academic publications from the *Journal of Information Technology and Tourism* and the proceedings of the Enter conferences. Overall, the use of these databases ensures the reliability, validity and timeliness of the articles retrieved (Law *et al.*, 2016; Gomezelj, 2016). Data used for this study were collected from January to March 2017, while the search was confined to the period of 2000-2016.

We adopted several search criteria. First, only full-length empirical and review/policy articles were included. Second, other articles, such as conference papers and book chapters, were excluded. Third, only empirical studies were considered in this study. Thereafter, the authors carefully read each selected article based on the aforementioned criteria and, thus, determined whether the article could be included in the analysis. We used different sets of keywords to build the target populations and samples. First, we searched the keywords "Business Intelligence" and "Big Data" on both databases. Second, we narrowed down our target population to look for work related to the hospitality and tourism areas by leveraging the following searches: matching "Business Intelligence" with the keywords "Hospitality" and "Tourism" separately and, similarly, matching "Big Data" with the keywords "Hospitality" and "Tourism" separately. A search using the keyword "Business Intelligence" in conjunction with other keywords representing BI components (such as "data warehouse", "data mining", etc.) in the titles, abstracts and keywords returned 70,212 and 42,158 academic publications on Scopus and Web of Science, respectively, over the period of 2000-2016. Apparently, there is a wide distribution over time and a linear and relevant growth over the past 16 years for Scopus-indexed works (Figure 1). The situation is significantly different for articles indexed in WOS, with low numbers (less than 10 articles per year), due to the lower coverage than Scopus.

A slightly different situation appears when focusing on BD. A search using the keyword "Big Data" in the titles, abstracts and keywords returned 29,101 and 18,159 academic publications on Scopus and Web of Science, respectively, over the period of 2000-2016. The time distribution is largely uneven and testifies to the very recent growth of interest in the topic. A look at Figure 2 shows an almost exponential growth, with an acceleration during the past five years.

Finally, we refined our search to ensure the direct relevance of the selected academic works to the wide area of tourism and hospitality. We have, therefore, constrained our search on publications including "travel, tourism, tourist, hospitality or leisure" in their titles, abstracts and keywords. Titles and abstracts have been manually inspected to further select work actually dealing with BI and BD. The consensus of multiple experienced researchers based in different academic institutions and countries and with dissimilar research and

cultural backgrounds and skills is thought to have minimized any potential personal bias during this last step of the selection process. The final data set, after checking for titles that were present in both databases (duplications), includes 77 articles related to BI and 96 articles related to BD. The final sample does not include duplicated articles. As mentioned, in the following section, we review the sample of articles, based on the following features: research topic; conceptual and theoretical characterization; sources of data; type of data and size; data collection methods; data analysis techniques; data reporting and visualization.

4. Results and discussion

The total number of BD articles in the area of tourism and hospitality selected as described above is 96. The number of BI publications is 77. Their time distribution is given in Figure 3 (for the past five years). Incredibly, it seems that besides the hype about the BD issue, not many hospitality and tourism researchers have decided to pay some effort in studying these topics so far, and only a handful of them have invested time and resources in considering the possibilities of an application of BD to the tourism and hospitality field.

What is more interesting is the fact that only 17 of these BD articles appear in tourism or hospitality journals, and, therefore, are more accessible to the tourism academic community. The same situation is found for the BI articles: only nine are available in tourism journals. All the others are found in computer science (mainly), transportation, marketing management or geography journals. In what follows, we describe the main features of a selection of articles whose topics are closely related to the domains of hospitality and

Figure 1.
Cumulative time
distribution of BI
works published over
the past 16 years in
Scopus and Web of
Science (WOS)

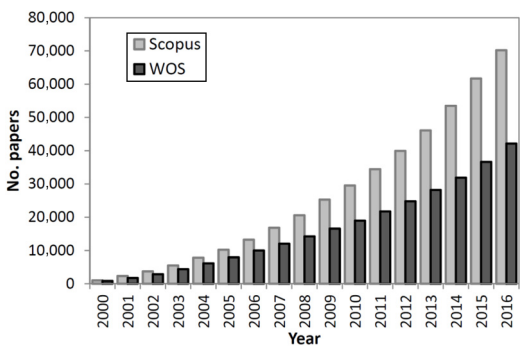
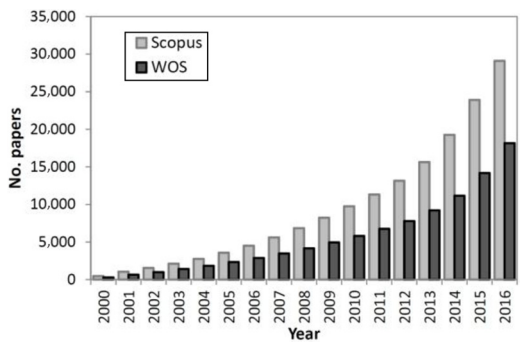


Figure 2.
Cumulative time
distribution of Big
Data works
published over the
past 16 years in
Scopus and Web of
Science (WOS)



tourism. These publications have appeared in both hospitality and tourism outlets and other scientific journals, mainly in the areas of computer science, transportation, marketing management and geography.

4.1 A critical discussion on the articles dealing with business intelligence

A selected number of BI articles with their main characteristics are reported in Table II.

The recurring terms in the titles and abstracts are summarized in the word cloud of Figure 4. A word cloud is the visual representation of the frequency with which words are found in a given context, providing a perceivable image of the most prominent terms and related themes (i.e. higher frequency of a term equals larger size). Here, most of the words are rather generic and show that the subject area has been treated through a relatively traditional approach.

The “tourism” BI literature focusses mainly on themes such as the organization of destination marketing information systems (Ritchie and Ritchie, 2002), methods for the analysis of specific tourists’ segments (Barbieri and Sotomayor, 2013), the examination of competitive intelligence practices in the hospitality sector (Köseoglu *et al.*, 2016) and frameworks for managing and analysing data (Fuchs *et al.*, 2013; Höpken *et al.*, 2015). Interestingly, among the most recent tourism BI publications, four (actually 22 per cent) are also catalogued in the BD listings (Fuchs *et al.*, 2014; Lam and McKercher, 2013; Marine-Roig and Anton Clavé, 2015; Qiao *et al.*, 2014). This is a clear indication of the fact that tourism scholars (at least those few who treat these topics) have well understood the capability of BD to provide insights that are useful for enriching the BI practices of destinations and tourism and hospitality operators.

From the 77 originally identified BI articles, 31 of them were selected for further investigation and critical discussion. These articles were selected as they deal with topics related to the domain of BI (Table II). First, only one article is purely conceptual by its nature. In this article, Pope *et al.* (2009) conceptualize the challenges and analytical opportunities found in collecting large volumes of data from airline websites and travel agencies. All other identified publications use either structured (i.e. sixteen papers), unstructured (i.e. six papers) or both types of data (i.e. eight papers). Second, two papers are conducting research about BI in the hospitality and tourism domain. More precisely, Ritchie and Ritchie (2002) analyse survey data to assess tourism industry stakeholders’ knowledge, needs and current use of research results and tools in the area of BI. Similarly, in a more recent publication, Köseoglu *et al.* (2016) evaluate awareness and knowledge about competitive intelligence in the hotel industry. Interestingly, the authors find that the most crucial competitive intelligence activities include price comparisons as well as the analysis

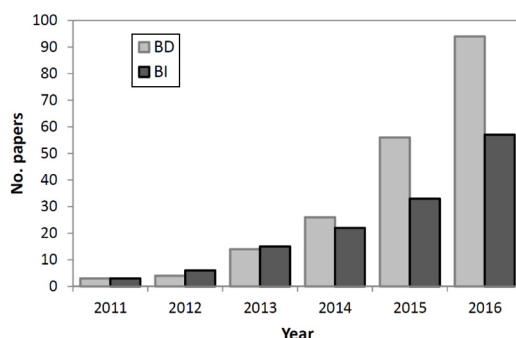


Figure 3.
Cumulative time
distribution of BD
and BI tourism and
hospitality related
works for the past
five years (base year
is 2011)

Table II.
BI works in
hospitality and
tourism (selected
works)

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Anadio and Procaccio (2016). Competitive analysis of online reviews using exploratory text mining	Text-based online review analysis using exploratory text mining techniques and visual analytics for SWOT analysis, applied to the hotel industry	Empirical	Online reviews from TripAdvisor	Unstructured	Manually (one-time)	LDA topic model, random forest classification	Dashboard, SWOT analysis
Archeitz et al. (2013). Web usage and content mining to extract knowledge for modelling the users of the Bidasoa Turismo website and to adapt it	Combined Web usage and content mining to generate user navigation profiles and semantically enriched user interest profiles as input to website optimization and marketing	Empirical	Web page content and Web server log files of Bidasoa Turismo website	Structured and unstructured	Automatically	PAM (Partitioning Around Medoids) clustering with Edit Distance sequence alignment method, SPADe (Sequential Pattern Discovery using Equivalence classes), LDA topic model	N/A
Ashtabor et al. (2007). Logit models for forecasting nationwide intercity travel demand in the USA	Nested mixed logit models to estimate market share of automobile and commercial air transportation	Empirical	American Travel Survey	Structured	One-time collection of existing secondary data	Nested and mixed logit models	Market share plots for five income groups
Carrasco et al. (2013). A multidimensional data model using the fuzzy model based on the semantic translation	Fuzzy model-based multidimensional data model to solve Opinion Aggregation when integrating heterogeneous information (including unstructured data)	Empirical	Web pages for data extraction are Atrapalo, Booking, eDreams, Expedia, TripAdvisor	Structured and unstructured	Automatically (periodically)	Explorative Data Analysis (EDA)	Dashboard, On-Line Analytical Processing (OLAP)
Chen and Tsai (2016). Data mining framework based on RST to improve location selection decisions: A case study of a restaurant chain	Data mining framework based on RST to support location selection decisions	Empirical	Survey of 33 directly managed stores of a restaurant chain	Structured	Manually	RST	N/A

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Chiang (2015). Applying data mining with a new model on CRM systems of online airlines and travel agencies to identify decision rules for discovering market segments	Mining high-value family travellers for CRM systems of online airlines and travel agencies to identify decision rules for discovering market segments	Empirical	Customer survey	Structured	Manually (one-time)	RFM model, Analytic hierarchy process (AHP), C5.0 decision tree	N/A
Dursun and Caber (2016). Using data-mining techniques for profiling profitable hotel customers: An application of RFM analysis	Profiling hotel customers by recency, frequency and monetary (RFM) indicators	Empirical	CRM system	Structured	Manually (one-time)	RFM model, self-organizing map (SOM), k-means clustering	Self-organizing map (SOM)
Fuchs <i>et al.</i> (2013). A knowledge destination framework for tourism sustainability: a BI application from Sweden	A destination management information system focusing on OLAP to measure proportion of tourists with the smallest ecological footprint	Empirical	Customer feedback data (survey-based)	Structured and unstructured	Manually (one-time) and automatically (periodically)	Explorative data analysis (EDA)	Html-based Web application, dashboards, OLAP
Fuchs <i>et al.</i> (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden	BI-based knowledge infrastructure implemented at the Swedish mountain destination, Åre and examples of use by tourism managers	Empirical	Web search, booking and feedback data (e.g. survey-based, user-generated content)	Structured and unstructured	Manually (one-time) and automatically (periodically)	Explorative Data Analysis (EDA) and machine learning (SVM, NB, Nearest Neighbour)	Html-based Web application, dashboards, OLAP
Holland <i>et al.</i> (2016). The role and impact of comparison websites on the consumer search process in the US and German airline markets	Examines how consumers search for airline tickets based on a comparative analysis of the US and German markets	Empirical	Click-stream panel data by ComScore	Structured	Manually (one-time)	Consideration set theory	N/A

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Hyphen et al. (2015) , BI for cross-process knowledge extraction at tourism destinations	A novel approach for BI-based cross-process knowledge extraction and decision support for tourism destinations	Empirical and conceptual	Web search, booking and feedback data (e.g. user-generated content)	Structured and unstructured	manually (one-time) and automatically (periodically)	Explorative data analysis (EDA) and data mining techniques (Decision Trees, Association Rule Mining); Multi-dimensional data warehouse modelling Back propagation neural networks	Html-based Web application, dashboards, OLAP
Hsieh (2011) , Employing a recommendation expert system based on mental accounting and artificial neural networks into mining BI for study abroad's P/S recommendations* Hsieh (2009) , Applying an expert system into constructing customer's value expansion and prediction model based on AI techniques in leisure industry	A recommendation Expert System for travel agencies based on mental accounting and artificial neural networks An Expert System platform addresses customer's value analysis based on artificial intelligence	Empirical Empirical	Online (student) survey about travel motivations and final decision making Online customer survey	Structured Structured	Manually (one-time) Automatically (periodically)	 Self-organizing feature map neural network for cluster analysis	N/A N/A
Kisilevich et al. (2013) , A GIS-based decision support system for hotel room rate estimation and temporal price prediction. The hotel brokers' context	A tool that assists travel intermediaries to acquire missing strategic information about hotels to leverage profitable deals. The GIS-based DSS estimates room rates using hotel and location characteristics	Empirical	OpenStreetMap data (public), Private data by a hotel brokerage, static names of hotels, internal IDs, location coordinates, hotel facilities, room amenities, hotel categories, dynamic room prices for one night customers received during their search, date of order	Structured	Automatically (periodically) and manually (one-time)	MDS; Voronoi tessellation partitioning; additive regression with isotonic regression; Locally Weighted Learning with Linear Regression; LibSVM nu-SVR; Multilayer Perceptron (ANN)	MDS component for exploratory data Analysis and Graphs to visualize price estimation results

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Kosoglu et al. (2016) , Competitive intelligence practices in hotels	Assessment of awareness and knowledge about competitive intelligence efforts in the hotel industry	Empirical	23 hoteliers' knowledge and awareness about competitive intelligence	Unstructured	In-depth interview	N/A	N/A
Kwok and Yu (2016) , Taxonomy of Facebook messages in business-to-consumer communications: What really works?	Combines machine learning and human intelligence to analyse Facebook messages initiated by hospitality companies	Empirical	2,654 Facebook messages initiated by 26 hospitality companies	Unstructured and structured	Automatically/ manually	Machine learning (SVMs)	Taxonomy of Facebook message types
Li et al. (2015) , Identifying emerging hotel preferences using emerging pattern mining technique	Identification of emergent hotel features by extracting frequent keywords from online reviews	Empirical	118,000 online reviews from TripAdvisor	Unstructured	Automatically (one-time)	Unsupervised feature extraction by frequent keywords, emerging pattern mining (EPM)	N/A
Lu and Zhang (2015) , Imputing trip purposes for long distance travel	Machine learning methods estimate trip purposes for long-distance passenger travel	Empirical	A passively collected long-distance trip data set is simulated from the 1996 American Travel Survey	Structured	Manually (one-time)	Decision tree and meta-learning	Confusion matrices from trip purpose imputation models
Marne-Roig and Anton Clavé (2015) , Tourism analytics with massive user-generated content: A case study of Barcelona	Studying online image of Barcelona as transmitted via social media through the analysis of more than 100,000 relevant travel blogs and online travel reviews	Empirical	Heterogeneous including the travel blogs, webpages, travelogues and travel reviews about Barcelona Heterogeneous data including the travel blogs, webpages, travelogues and travel reviews about Barcelona (250,000 pages)	Unstructured	Data was extracted automatically through Offline Explorer Enterprise	Pre-processing: web content mining, language detection, user's hometown, cleaning, debugging, processing: parser settings and categorizations through Site Content Analyzer	Tables created through word count

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Pope et al. (2009) . Conceptual framework for collecting online airline pricing data: Challenges, opportunities, and preliminary results	Challenges and opportunities to collect large volumes of data from airline websites and travel agencies are discussed. Research questions are highlighted that can be investigated with this type of data	Conceptual	N/A	N/A	N/A	N/A	N/A
Ritchie and Ritchie (2002) . A framework for an industry supported destination marketing information system	Guidelines for the establishment of a comprehensive destination marketing information system (DMIS)	Empirical	Industry stakeholders' knowledge needs and current use of research and intelligence (Inter-National Travel Survey)	Primary survey data from 68 individuals Secondary data (travel surveys)	Semi-structured interview	N/A	N/A
Rossetti et al. (2016) . Analyzing user reviews in tourism with topic models	A description of the topic model method with application focus on the tourism domain	Empirical	Yelp Data set Challenge; TripAdvisor Dataset	Structured and unstructured	Yelp is existing data set; TripAdvisor automatically collected by crawler	KNN-UB, KNN-IB, Probabilistic Matrix Factorization (PMF)	Illustrative examples for selected topics related to multi-criteria dimensions
Sanchez-Franco et al. (2016) . Online Customer Service Reviews in Urban Hotels: A Data Mining Approach	Extraction of features from hotel reviews and analysis of their relationship with guests' hotel rating in the online travel agencies environment	Empirical	19,318 hotel reviews from 2014 to 2015 from booking.com	Structured and unstructured	Automatically (one-time)	Pathfinder network scaling, principal component analysis (PCA), linear mixed-effects regression	N/A
Snaveley et al. (2008) . Modeling the world from internet photo collections	Presents algorithms and results as a step towards 3D modelling of the world's well-photographed sites, cities, and landscapes from internet imagery	Empirical	Flickr	Large sets of image data	Automatically downloaded from Flickr	Keypoint detection scale invariant feature transform (SIFT) keypoint detector and matching (by approximate nearest neighbors (ANN) kd-tree); Structure for motion (to recover camera parameters);	reconstructed scenes and photo connectivity graphs for 11 sites

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Solnet et al. (2016) . An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue	Market Basket Analysis to identify and predict the purchasing behavior of customers based on their expenditure patterns to determine the most attractive additional products and services	Empirical	56,906 guest sales records from a luxury hotel group in Australia from 2009 to 2014	Structured	Automatically (one-time)	geo-registration (by digital elevation maps) Multivariate logit model	N/A
Sun et al. (2016) . Chinese Customers' Evaluation of Travel Website Quality: A Decision-Tree Analysis	Identification of critical attributes that influence quality levels of a customer's travel agency's website experience	Empirical	Survey data from 25 individuals	Structured	Manually	Attention: interest-desire-action (AIDA) model, C4.5 decision tree	N/A
Tseng and Won (2016) . Integrating multiple recommendation schemes for designing sales force support system: A travel agency example	Proposes a design of sales force support (SFS) system with BI methodologies	Empirical	N/A	N/A	N/A	Explorative data analysis (EDA) and data mining (e.g. sequential pattern discovery)	Dashboards, OLAP
Versichele et al. (2014) . Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent	Spatiotemporal tourism behaviour by mining of association rules in tourist attraction visits based on Bluetooth tracking methodology	Empirical	17,496 Bluetooth devices being detected over 235,397 time intervals by 15 Bluetooth sensors in Ghent in 2015	Structured	Automatically	<i>A priori</i> association rule mining	Visit pattern maps

(continued)

Table II.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Source(s) of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Wu <i>et al.</i> (2010) , Data mining for hotel occupancy rate: An independent component analysis approach	Identification of major factors determining the hotel occupancy rate and incorporation of these factors to decompose hotel occupancy rates and examine the effect of each factor on the hotel occupancy rate	Empirical	Monthly hotel occupancy rate time series for each district of Hong Kong from January 1996 to May 2009	Structured	Manually	Independent component analysis (ICA)	N/A
Zhang and Huang (2015) , Mining tourist motive for marketing development via twice-learning	Application of twice-learning framework to predict tourists' travel motives from tourists' external and internal features, useful for targeted marketing strategy development	Empirical	On-site surveys in Nanjing, China, from October to November 2012 with 121 responses	Structured	Manually	Twice-learning framework, neural networks, C4.5 decision tree, NB	N/A
Zhu <i>et al.</i> (2016) , Get into the spirit of a location by mining user-generated travelogues	Location information extraction from user-generated travelogues, examining contents and structures of travelogues, as well as their interplay	Empirical	80,384 travelogues related to tourist destinations in the USA	Unstructured	Manually	Gazetteer-based location detection, semantic correlation detection by natural language parsing techniques, location concept network by PLSA	N/A

When it comes to the aggregation and (sentiment) analysis of user-generated content, [Rossetti *et al.* \(2016\)](#) apply both K-nearest neighbour user based (KNN-UB), K-nearest



Figure 4.
Word cloud with the
most used terms in
the BI papers selected

neighbour item based (KNN-IB) and probabilistic matrix factorization (PMF) techniques. By contrast, [Marine-Roig and Anton Clavé \(2015\)](#) use parsing and categorizations through a word-frequency-based site content analyzer. Moreover, to analyse user-generated content, [Fuchs et al. \(2014\)](#) use ML techniques, like support vector machines (SVM), Naïve Bayes (NB) and K-Nearest Neighbour (KNN). The work by [Kwok and Yu \(2016\)](#) combines ML and human intelligence to analyse Facebook messages initiated by hospitality companies. More precisely, the authors use SVMs to classify Facebook messages as for instance “popular/less popular” and different message types, to identify relevant keywords to define a taxonomy of Facebook messages ([Kwok and Yu, 2016](#)). Moreover, some authors use Latent Dirichlet Allocation (LDA) topic models to analyse online reviews from TripAdvisor ([Amadio and Procaccino, 2016](#)) and web content in conjunction with web server log file data of a tourism DMO platform ([Arbelaitz et al., 2013](#)). The article by [Li et al. \(2015\)](#) applies feature extraction by frequent keywords and Emerging Pattern Mining (EPM) techniques for identifying emergent hotel features based on 118,000 online reviews from TripAdvisor. In a similar vein of analysis, [Sánchez-Franco et al. \(2016\)](#) extract features from hotel reviews by using pathfinder network scaling (PNS), principal component analysis (PCA) and linear mixed-effects regression. Finally, [Zhu et al., \(2016\)](#) apply Gazetteer-based location detection and semantic correlation detection by natural language parsing techniques to extract location information from travelogues.

Artificial neural networks (ANN) are used in the form of self-organizing feature maps as a means of clustering large amounts of data ([Hsieh, 2009](#); [Dursun and Caber, 2016](#)), or in the form of back-propagation networks ([Hsieh, 2011](#); [Zhang and Huang, 2015](#)) and multilayer perceptrons ([Kisilevich et al., 2013](#)) as classification techniques. Moreover, the article by [Chen and Tsai \(2016\)](#) applies rough set theory (RST) to support customers’ location decisions in the context of a restaurant chain.

Finally, the BI papers identified in the hospitality and tourism domain also comprise (multi-variate) statistical techniques, such as multi-dimensional scaling (MDS) ([Kisilevich et al., 2013](#)), mixed logit models ([Ashiabor et al., 2007](#); [Solnet et al., 2016](#)) or mathematical modelling ([Holland et al., 2016](#)). When it comes to 3D modelling from internet imagery of photographed sites ([Snaveley et al., 2008](#)), very specialized techniques, such as scale invariant feature transform (SIFT) key-point detectors, approximate nearest neighbors and *kd*-tree analysis is applied. Similarly, [Kisilevich et al. \(2013\)](#) apply highly specialized techniques, such as Voronoi Tessellation Partitioning, additive regression with isotonic regression, locally weighted learning and LibSVM nu-SVR.

Concerning data collection methods, twelve articles feature a one-time manual approach, while nine others feature an automated periodical approach for collecting data. Four papers display a combination of both methods for data collection. When it comes to visualization of BI-based knowledge, eight articles feature dashboards and/or OLAP ([Kisilevich et al., 2013](#)). [Ashiabor et al. \(2007\)](#) develop special visualization forms, such as market share plots. Likewise, [Snaveley et al. \(2008\)](#) reconstruct scenes and photo connectivity graphs for 11 well-photographed sites.

To conclude, in relation to research type and topic, data source and structure, data collection methods as well as data analysis and visualization techniques, the discussed BI papers show a broad variety and, by that, clearly demonstrate the diversity and multi-faceted dimensions of the domain of BI. In numbers, while 10 of the 31 analysed publications that deal with BI applications relate to tourism destinations (i.e. local, regional or national tourism organization, destination management organization), 11 are linked to hospitality businesses. Similarly, nine papers can be assigned to tourism and travel industries (i.e. five BI studies relate to travel agencies and four to the airline business). Finally, only one BI

study identified by our review has been conducted in the field of leisure. Nevertheless, all analysed papers on BI applications in travel, tourism and hospitality clearly underpin the huge potential for future applications of BI in this domain.

4.2 A critical discussion on the articles dealing with big data

A selected number of BD articles with their main characteristics are reported in [Table III](#).

A detailed reading of the *tourism* abstracts highlights some other interesting facts.

The first thing to notice, for what concerns BD, is that the abstracts and titles contain rather *generic* terms, with no or little reference to the specific terminology often used in works about BD ([Figure 5](#)).

Some articles feature a rather general and somewhat conceptual discussion about BD or about the general importance of using BD for improving and extending present research activities ([Buhalis and Foerste, 2015](#); [Dolnicar and Ring, 2014](#); [Wang et al., 2015](#)). Despite the call for a better integration between official statistics and BD ([Heerschap et al., 2014](#); [Lam and McKercher, 2013](#)), not many of the identified publications attempt to find a solution. For instance, [Yang et al. \(2014\)](#) use Web traffic volume data of a destination marketing organization to predict hotel demand, showing an improvement in the error reduction in contrast to more traditional forecasting models. Similarly, [Önder et al. \(2014\)](#) use Flickr geotagged photos to assess the presence of tourists in Austria, showing that the method provides more reliable outcomes for cities than at a regional level. Finally, [Fuchs et al. \(2014\)](#) demonstrate how BD analytics can be beneficial for BI practices in a tourism destination and propose an architectural solution that combines different sources of data, such as customers' Web search, booking and feedback data.

Advanced approaches, such as ML techniques, AI or Bayesian classification methods are practically ignored, and the most used technique is a simple statistical textual analysis of pieces collected online from which the authors derive a number of insights (see [Berezina et al., 2016](#) or [Lu and Stepchenkova, 2015](#)). A notable exception are the papers by [Menner et al. \(2016\)](#) and [Schmunk et al. \(2014\)](#) that perform sentiment analysis on a large corpus of user generated contents by using advanced techniques, such as SVMs, NB classifiers and latent semantic indexing.

Not many other articles actually use online sources. An exception is the article of [Xiang et al. \(2015\)](#). The authors analyse a large corpus of tourists' reviews to derive a number of interesting considerations about hotel guest experience and its association with satisfaction ratings ([Xiang et al., 2015](#)). Similarly, [Marine-Roig and Anton Clavé \(2015\)](#) collect a large quantity of user-generated comments (i.e. travel blogs and online travel reviews) concerning the area of Barcelona and deduce the perceived image of the city through these reports. Along this line, [Park et al., 2016](#) analyse tweets generated by cruise travellers showing their main interests and preferences, thus providing useful suggestions for feasible marketing strategies. [Mariani et al. \(2016\)](#) examine Facebook pages of Italian destinations revealing how destinations use the social platform and which posts' characteristics have the biggest impact for actively engaging visitors. Finally, [d'Amore et al. \(2015\)](#) present a hard- and software system for helping in the troublesome collection of data from online social media platforms. Other types of records are even more sparingly used. Examples are [Kasahara et al. \(2015\)](#), who study GPS tracks and a possible method for inferring transportation modes, or [Gong et al. \(2016\)](#), who use taxi trajectory data (GPS-based) for guessing the probability of points of interest to be visited in a city, thus deducing possible trip purposes and travel patterns. It must be noted that all these publications use relatively *small* quantities of data (in the range of a few dozen thousand records) compared with what would be (probably) available for the studies. Only a few publications, in fact, use relatively large

Table III.
BD works in
hospitality
and tourism
(selected works; in
“type of data and
size” an asterisk
indicates large
quantities of data,
> 100 000 records)

Article (author and title)	Research topic	Type of paper (conceptual/ empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Bahalis and Foerste (2015) . SoCoMo marketing for travel and tourism: Empowering co-creation of value	Proposes social context mobile (SoCoMo) marketing as a new framework that enables marketers to increase value for all stakeholders at the destination	Conceptual	N/A	N/A	N/A	N/A	N/A
Carter (2016) . Where are the enslaved?: TripAdvisor and the narrative landscapes of southern plantation museums	Explores what visitors learn about the history of the enslaved on two tours (Laura and Oak Valley) and how they participate in the narrative construction of the plantation	Empirical	TripAdvisor visitor reviews (Laura and the Oak Alley museums, USA)	Unstructured	Web (manual) scraping	Word frequency and words associations in reviews	Standard tables
Dolnicar and Ring (2014) . Tourism marketing research: Past, present and future	Critical review of tourism marketing research	Literature review	N/A	N/A	N/A	N/A	N/A
Fuchs et al. (2014) . BD analytics for knowledge generation in tourism destinations - A case from Sweden	Presents a knowledge infrastructure implemented at the Swedish mountain tourism destination, Are and examples of use by tourism managers	Empirical	Web search, booking and feedback data (e.g., survey-based, user-generated content)	Structured and unstructured	Data Warehouse (DW) including Facts and Dimensions Tables	On-Line Analytical Processing (OLAP); SVM, NB and KNN	Html-based web application
García-Pablos et al. (2016) . Automatic analysis of textual hotel reviews	Describes OpenNER, a NLP platform applied to the hospitality domain to automatically process customer-generated textual content	Empirical	Online reviews from Zoover and HolidayCheck	Unstructured	Web crawler	Natural Language Processing; Named Entity Recognition, Sentiment Analysis and Opinion Mining	Standard tables
García-Palomares et al. (2015) . Identification of tourist hot spots based on social networks: A comparative analysis of european metropolises using photo-sharing services and GIS	Use of photo-sharing services for identifying and analyzing the main tourist attractions in eight major European cities	Empirical	Panoramio photos	Unstructured (*)	Panoramio website API + ArcGIS	Density graphs, spatial autocorrelation	Standard tables and Anselin Local Moran's I graph
Gretzel et al. (2015) . Smart tourism: Foundations and developments	Defines smart tourism, sheds light on current smart tourism trends, and lays out its technological and business foundations	Conceptual	N/A	N/A	N/A	N/A	N/A

(continued)

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Gunter and Oeder (2016) . Forecasting city arrivals with Google analytics	10 Google Analytics website traffic indicators from the Viennese DMO website are used to predict actual tourist arrivals to Vienna	Empirical	Google analytics variables collected on a monthly basis over the period August 2008–October 2014	Structured	Simple access to Google analytics	VAR model class: BVAR, FAVAR, BF-FAVAR	Basic tables of descriptive statistics
He et al. (2016) . Travel-package recommendations leveraging social influence of different relationship types	Develops a probabilistic topic model leveraging individual travel history and social influence of co-travellers to capture personal interests and propose a recommendation method to use the proposed model	Empirical	Structured travel records on travel packages	Structured	Access to a private company database	Biggs sampling	Basic tables of descriptive statistics
Jackson (2016) . Prediction, explanation and bigger data: A middle way to measuring and modelling the perceived success of a volunteer tourism sustainability campaign based on 'nudging'	Uses "automatic linear modelling" that can cope with BD and presents the results as visualizations	Empirical	Structured (responses from questionnaire)	Structured	Survey	Automatic linear modelling and preparation through IBM SPSS	Basic tables of descriptive statistics and graphs stemming from automatic linear modelling (IBM SPSS)
Kong and Song (2016) . A study on customer feedback of tourism service using social BD	Design of an analysis model for the top Korean travel agency to help the company improve customer satisfaction and service quality	Empirical	Internal sources (emails, counselling data, bulletin information, after use comments/evaluations) and external sources (Twitter, Facebook, OnlineNews, Blog, Community)	Mostly unstructured (e.g. data from emails, social media networks) and several structured (e.g. bulleting info)	BuzzMonitoring (Types and proportion of keywords from the extracted data are digitized to analyse incidents and phenomena)	BuzzMonitoring including the following modules: NLP, data clustering, text summarization, sentiment analysis, structure data joiner	No table nor graphs stemming from the Buzz Monitoring
Law et al. (2011) . Identifying changes and trends in Hong Kong outbound tourism	Trends in Hong Kong outbound tourism in terms of Future trip intentions Travel destinations Motivation to travel	Empirical	Tourism behaviour survey data	Survey questionnaire	Historical domestic Surveys	Data mining, association rules, contrast set mining	Tables

(continued)

Table III.

Table III.

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Mariani <i>et al.</i> (2016). Facebook as a destination marketing tool: Evidence from Italian regional destination management organizations	Explores how Italian regional destination management organizations (DMOs) strategically employ Facebook to promote and market their destinations, and improves on the current metrics for capturing user engagement	Empirical	Overall number of Facebook posts posted on the official Italian regional DMOs' Facebook pages	Structured (*)	Data extractor based on Facebook APIs	Data parser and analyzer calculating per post statistics	Tables created through data analyzer module. Graphs created through the data visualizer module
Marine-Roig and Anton Clave (2015). Tourism analytics with massive user-generated content: A case study of Barcelona	Studying the online image of Barcelona as transmitted via social media through the analysis of more than 100,000 relevant travel blogs and online travel reviews (OTRs) written in English	Empirical	Heterogeneous including the travel blogs, webpages, travelogues and travel reviews about Barcelona (250,000 pages)	Unstructured (*)	Data were extracted through Offline Explorer Enterprise (OEE)	Pre-processing: web content mining, language detection, user's hometown, cleaning, debugging. Processing: parser settings and categorizations through Site Content Analyzer Grey 2-tuple linguistic evaluation (expert evaluations)	Tables created through word count
Mi <i>et al.</i> (2014). A new method for evaluating tour online review based on grey 2-tuple linguistic	Evaluation of online reviews	Empirical	Reviews from tourism website	Unstructured	Web crawler		Standard tables
Mocanu <i>et al.</i> (2013). The twitter of babel: Mapping world languages through microblogging platforms	Survey on worldwide linguistic indicators and trends through	Empirical	Large-scale data set of geotagged tweets	Structured and unstructured (*)	Twitter API	Language detection (Google Chromium Compact Language Detector) - Geographical analyses Clustering analysis	Maps + charts and tables
Noguchi <i>et al.</i> (2016). Advanced, high-performance BD technology and trial validation	Presents technology for analyzing data and location data	Conceptual, application design, case study	Smartphone application	Structured: app's usage logs, location data and individual	From app logs		Maps + charts and tables
Orellana <i>et al.</i> (2012). Exploring visitor movement patterns in natural recreational areas	Explores the properties of the collective movement of visitors in recreational natural areas	Empirical	Global Positioning System tracking data	Structured (GPS data)	Recording of GPS data	Kernel-density function classification and Generalized Sequential Patterns	GIS + standard tables

(continued)

Article (author and title)	Research topic	Type of paper (conceptual/ empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Paldino <i>et al.</i> (2015) . Urban magnetism through the lens of geo-tagged photography	Tastes of individuals, and what attracts them to live in a particular city or spend vacation there	Empirical	Geo-tagged photos	Structured: metadata (*)	Flickr API	Identification of resident, tourist and unknown, statistical analysis, network analysis (origin/ destination)	Maps + charts and tables
Park <i>et al.</i> (2016) . Using twitter data for cruise tourism marketing and research	Analysis social media data on cruise tourism	Empirical	Tweets containing search words	Unstructured/ structured	Twitter API and Web scrapping	Word frequency, content analysis, and network analysis	Charts, tables, network diagrams
Raun <i>et al.</i> (2016) . Measuring tourism destinations using mobile tracking data	Measure space-time tracking data to analyse, monitor and compare destinations based on data describing actual visits	Empirical	Anonymized roaming data of the foreign mobile phone call detail records	Structured	From telcom operator	Statistical analyses, ArcGIS for spatial analyses, binary logistic regression	Charts, tables, maps
Shi <i>et al.</i> (2016) . Applying semantic web and BD techniques to help collect data, and implement balance model referring to stakeholders of tourism intangible cultural heritage	Apply semantic web and BD techniques to help collect data, and implement platform and questionnaire design to construct stakeholder balance model for tourism intangible cultural heritage	Empirical	Questionnaire + User Generated Content (reviews)	Structured/ unstructured	Unknown UGC	Structural equation model, path analysis	Charts, tables
Su <i>et al.</i> (2016) . Characterizing geographical preferences of international tourists and the local influential factors in china using geo-tagged photos on social media	Characterize geographical preferences of international tourists and quantify local influential factors of tourists' destination preferences across time and space and origins	Empirical	Metadata online geotagged photos	Structured	Flickr API	Statistical and spatial analyses	Charts, tables, maps

(continued)

Table III.

Table III.

Article (author and title)	Research topic	Type of paper (conceptual/ empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Sun <i>et al.</i> (2016). Internet of things and BD analytics for smart and connected communities	Integration of Internet of Things (IoT) and BD analytics for smart connected communities	Conceptual + case study	Design of an IoT system, personal sensors, open data, and participatory sensing to enhance the services in the area of tourism and cultural heritage with a Context-Aware Recommendation System	N/A	N/A	N/A	N/A
Supak <i>et al.</i> (2015). Geospatial analytics for federally managed tourism destinations and their demand markets	Examine the general geospatial demand for overnight recreation on federal lands prior to the 2008 recession and the specific geospatial demand for national park regions	Empirical	National Recreation Reservation Service reservations database	Structured (*)	DB access	Statistical and spatial analyses	Charts, tables, maps
Tang and Li (2016). Spatial network of urban tourist flow in Xi'an based on microblog BD	Study related to spatial network of tourist flows and its structure in the urban areas	Empirical	Geotagged microblog posts	Structured (*)	API from Sina Microblog	Statistical, spatial and network analyses	Charts, tables, maps, network diagrams
Wang <i>et al.</i> (2015). Revenue management: Progress, challenges, and research prospects	Discuss evolution and future developments of revenue management and use of BD analytics	Conceptual	N/A	N/A	N/A	N/A	N/A
Wood <i>et al.</i> (2013). Using social media to quantify nature-based tourism and recreation	Online posted photos are used to estimate visitation rates and travelers' origins. Compare to empirical data showing that crowd-sourced information can serve as reliable proxy for empirical visitation rates	Empirical	Empirical data sets that quantified visitation to 836 sites in 31 countries around the world + Flickr metadata	Structured (*)	Data set + Flickr API	Statistical and spatial analyse	Charts, tables, maps

(continued)

Article (author and title)	Research topic	Type of paper (conceptual/empirical)	Sources of data	Type of data and size	Data collection methods	Data analysis techniques	Data reporting and visualization
Yang <i>et al.</i> (2014). Predicting hotel demand using destination marketing organization's Web traffic data	Demonstrate the value of website traffic data in predicting demand for hotel rooms at a destination, and potentially future revenue and performance	Empirical	Website traffic data and local hotel room demand data	Structured	Google analytics + standard data	Statistical and time series forecasts	Charts, tables
Yang <i>et al.</i> (2016). The BD analysis of land use evolution and its ecological security responses in silver beach of china by the clustering of spatial patterns	Use remote-sensing images to analyse the land use evolution and to evaluate its ecological security	Empirical	Landsat satellite high-definition images	Pictures + metadata	Landsat DB (*)	Land use temporal statistical analysis	Charts, tables, maps

Table III.

Two specific papers are worth mentioning here. One is the study of global mobility of people conducted by [Hawelka et al. \(2014\)](#) who geotagged one year worth of tweets (almost one billion), deriving patterns and some characteristics of the movements of international travellers. The second is the report by [RocaSalvatella \(2014\)](#) who collected mobile phone traffic and credit card transactions data in Madrid and Barcelona for one month (about 700,000 phones and 170,000 cards), retrieving information about a number of detailed activities and expenditures of international visitors to the two cities. While 25 of the 96 analysed articles using BD applications are related to tourism destinations (i.e. local, regional or national tourism organization, destination management organization), 22 are linked to hospitality businesses. Similarly, 48 articles can be assigned to the travel industry. Finally, only one study identified by our review has been conducted in the field of leisure. Clearly, in cases where an article addressed both hospitality and destination aspects, such as with [Yang et al. \(2014\)](#), we had to make a choice on how to categorize it to avoid duplications; the criterion adopted was based on the prevalence of a theme over the other. Overall, the analysed papers leveraging on BD in travel, tourism and hospitality clearly underpin the huge potential for future applications of BD with and increasing use of BD stemming from online consumer reviews in the hospitality sector.

5.1 Conclusions

The systematic literature review reveals that there is an increase in hospitality and tourism management literature applying analytical techniques to large quantities of data. However, this research field is fairly fragmented in scope and limited in methodologies and displays several gaps. A conceptual framework that would help identify critical business problems and link the domains of BI and BD to tourism and hospitality management and development is missing. Moreover, epistemological dilemmas and consequences for theory development of BD-driven knowledge are still a *terra incognita*.



Figure 5.
Word cloud with the
most used terms in
the BD papers
selected

Before we introduce the theoretical and practical implications of our review, we briefly sketch a number of domain-independent challenges stemming from the implementation of a BI and BD environment (Jannach, 2016, p. 109). Here, we must also remark that there is a quite strong interplay between BI and BD. After all, data are the underlying resource for BI, and the relationships existing between the two are often so close that it is difficult to separate them, as already noted here and in other publications (Fan *et al.*, 2015; Lycett, 2013). BD has emerged as a disruptive technology with effects that will surely be reshaping BI, a domain completely relying on data analytics for the purpose of better decision-making.

In our survey, one important point concerns questions of how to collect automatically data represented in non-standardized formats, and where to store these huge volumes of data for later processing. In addition, combining different heterogeneous data sources is puzzling, as issues of data integration and data quality arise if data come from sources both inside and outside the organization. Probably, the most difficult question is how to extract “useful” knowledge from data to support better decision-making. For instance, richer data visualizations (e.g. through customizable dashboards) might be appropriate in one case, while for other cases more sophisticated prescriptive ML models might be more suitable for knowledge extraction. Finally, existing business models need to be adapted based on new insights gained from BI and BD. This similarly challenging task requires out-of-the-box thinking as well as cross-departmental work. Furthermore, it will require that academic and research institutions set up collaborations and synergies with industry stakeholders to get access to previously non-accessible data (Klein and Jacobs, 2016).

In recent years, more and more data have become available through the extensive use of online applications, data that are increasingly used for the analysis of consumer behaviours, the elaboration of marketing strategies, predicting trends, detecting frauds and for producing new, faster and more detailed statistics (Gandomi and Haider, 2015). The latter point is of utmost importance for the hospitality and tourism domains as most diagnostics and predictive tourism activities are based on official records provided by national or local statistical organizations, or are based on surveys conducted in “traditional” ways. However, the reliability of official tourism data has been questioned several times in the literature (Lam and McKercher, 2013). The reasons behind this questioning include the poor harmonization of data collection methods, the currency of data and statistical estimation procedures. Even more important is the issue that, with the growth of multiple forms of travels and stays, many visitors go unobserved (Baggio, 2016). Thus, one possible solution to improve data quality is resorting to the records of innumerable trails that millions of individuals leave online when using the many currently available technological platforms (Baggio, 2016). Indeed, *electronic footprints* have shown to be a valuable source to assess travellers’ and tourists’ behaviour and related decision-making and knowledge sharing (Lu and Stepchenkova, 2015).

However, the attempt to apply knowledge created in *real time* on the base of tourists’ on-site behaviour is especially challenging, as, currently, BI applications are still a rarity in tourism (Fuchs *et al.*, 2014; Höpken *et al.*, 2015; Yuan and Ho, 2015). Consequently, real-time knowledge generation in the hospitality and tourism domain needs to be significantly improved through ubiquitous (i.e. mobile) end-user applications (Kolas *et al.*, 2015), showing the capacity to collect tourists’ real-time feedback (Kolas *et al.*, 2015) and to trace movement patterns most effectively (Shoval and Isaacson, 2010; Zanker *et al.*, 2010; Höpken *et al.*, 2012). Supplier-based knowledge sources from the digital destination eco-system (Baggio and Del Chiappa, 2014) can also be integrated in real-time, such as through product-profiles and available information automatically extracted from supplier websites and databases. Thus, real-time knowledge about suppliers’ service potential (property status), the complementarity of destination offers

and its evaluation through tourists' real-time feedback, all significantly support dynamic need fulfilment in a collaborative tourism destination environment (Fuchs *et al.*, 2016). To conclude, BD is mainly a collection of data generated from people, companies, groups and networks, implying that international and cultural differences still have a dominant influence (Klein and Jacobs, 2016). This, however, makes consolidation and interpretation of data, as well as their patterns and correlations on various aggregation levels, a highly complex and difficult task.

5.2 Theoretical implications and research needs

Though some researchers are already claiming "the end of theory" as the "data deluge makes the scientific method obsolete" (Andersson, 2008), there is still a strong need for theories addressing the consolidation and interpretation of data. More precisely, it is only a "theory", which is capable to deduce conclusions from (i.e. causal) patterns in data in a self-consistent way (Han, 2015). From an epistemological perspective, without theories, BD merely creates (algorithmically) generated numbers totally uncoupled from social realities. According to the German philosopher Hegel (1830), all the reasonable is *conclusion*. Thus, theories are and will always be required as the "narrative way" behind knowledge generation. From this perspective, Andersson (2008) claim of the "end of theory" would imply the end of mind (Han, 2015). Hence, mixed-method approaches, triangulation and sense making through various theoretical frameworks are still helpful to understand the broad landscape of BD. Moreover, BD might be complemented and combined with small data stemming from traditional data collection and analysis techniques, a reflection of what is happening right now (Kitchin and Lauriault, 2015), also in light of the difficulties that small and medium firms can face in equipping themselves with the right human resources, skills and capabilities (Coleman *et al.*, 2016). Obviously, theories are constantly revised and new concepts (e.g. long tail) developed in light of new empirical evidence. For instance, by referring to the above-sketched example on mobile customer relationship management in a tourism destination context (Kolas *et al.*, 2015), theories on consumer (i.e. tourist) behaviour help defining (i.e. modelling) the relevant tourism domain (e.g. tourists' decision-making during a stay in a typical winter destination). By doing so, the relevance of major data sources can be assessed. Thus, only those customer-based data will be collected (e.g. in real-time) which have an expected influence on desirable, ideally sustainable, consumer behaviour (Fuchs *et al.*, 2014; Höpken *et al.*, 2015). In addition, this leads to a systematic avoidance of a data overload.

Certainly, there are several knowledge-gaps and development needs in the domain of BI and BD in hospitality and tourism. As far as knowledge gaps are concerned, by referring to the relative share of BD and BI articles falling into the tourism and hospitality domains, one can conclude that tourism scholars are increasingly aware about the significance and potential impact of BD and BI on these business and societal domains. Nevertheless, the absolute number of articles published during the period under analysis was marginal. Thus, future work is needed to conceptualize and (e.g. prototypically) implement innovative BI solutions as well as to critically assess the use and usefulness of these BI applications and BD in the tourism and hospitality domains. Second, we lack a conceptual framework that helps to identify critical business problems in the hospitality and tourism domain (Xiang, 2016, p. 127). For example, the linkages between data analytics and smart tourism development are yet to be established (Gretzel *et al.*, 2015). Third, and related to the previous point, several issues related to BD, such as the epistemological dilemmas and consequences for theory development of data-driven knowledge, are still a *terra incognita* (Ekbia *et al.*, 2015). Fourth, more attention should be paid to issues problematizing critically the use of BD stemming from online platforms as a data source (Mariani and Borghi, 2018) and the role of

submission devices (Mariani *et al.*, 2019). For instance, to date most of the studies relying on online travel review platforms have not dug in depth about the differences in the way data are produced, frequently neglecting an accurate process of data understanding and data cleaning. Fifth, as discussed above, at the destination level, methodological work is needed to complement official statistics that keep their basic validity with the large quantity of data available online (Fan *et al.*, 2014; Kitchin and Lauriault, 2015). At the individual firm level, an important challenge will be to complement small data (SD), collected through traditional methodologies (such as traditional customer surveys), with BD stemming from online records. Thus, we envision that companies in hospitality and tourism will keep on conducting *ad hoc* traditional customer surveys, but at the same time should (and will) juxtapose the perception metrics stemming from SD with the actual behavioural data stemming from BD generated from online records (Heerschap *et al.*, 2014). In other words, the “variety” feature characterizing BD should be applied to data *tout court*, regardless of its source (e.g. traditional customer surveys vs online review platforms). Bringing BD and SD together might enrich the managerial insights of destination and hospitality marketers as SD typically relate to customer perceptions, while BD have the advantage of including also behavioural analytics. Finally, it seems that while management and tourism management scholars are becoming increasingly aware of the relevance of BD for BI, still their collaborations with computer and data scientists are rather episodic and related to specific types of work and research. However, there is a need to integrate progressively management and data science (Gerard *et al.*, 2016) in a more systematic way through the creation of multi- and inter-disciplinary research teams involving hospitality and tourism management scholars on the one hand and computer and data scientists on the other hand (Fuchs *et al.*, 2014; Höpken *et al.*, 2015). This might ensure that a diverse set of competences, capabilities and skills will be used to face a range of complex research issues with the right analytical tools.

5.3 Practical implications

After the identification of robust and validated methodologies, the development of hardware infrastructures and open software applications (Alaei *et al.*, 2017; Kirilenko *et al.*, 2017; Höpken *et al.*, 2017; Dergiades *et al.*, 2018) could provide researchers and practitioners with adequate tools for dealing with BD in tourism and hospitality settings. As highlighted, real-time BI delivers information about business processes as they occur. The inherent automated analysis capacity enables the initiation of corrective actions and the adjustment of business rules, to optimize business processes in real-time. One example would be mobile customer relationship management (m-CRM) applications, which automatically detect customer opportunities immediately communicated to customers’ smartphones. Thus, a promising task for BI in hospitality and tourism is to create such real-time data, which is currently unavailable, to reflect tourists’ on-site behaviour by means of ubiquitous (mobile) e-CRM applications (Sinisalo *et al.*, 2007; Vogt, 2011; Wang *et al.*, 2012; Kolas *et al.*, 2015). On the base of this new type of customer data, valuable knowledge for businesses and destination management can emerge through methods of real-time BI and data mining. Finally, real-time knowledge needs can be used to adapt intelligent ubiquitous (mobile) customer applications, thereby enhancing the match between customer needs and offered destination products. For instance, real-time travel patterns, characterized by *ad hoc* and less systematic travel decisions and activities, can be considered in activity-based transportation models and recommender systems (Hermans and Liu, 2016). For this purpose, the use of mobile phone data and credit-card data offers valuable insights into travellers’ activities and travel behaviour in real-time (Liu *et al.*, 2013). The aforementioned

trends are evolving fast and future developments in the wider field of AI applied to data mining and predictive learning look quite promising to enhance the intelligence capabilities of tourism and hospitality organizations and assist them in understanding fast changing and hypercompetitive markets that can translate into sustained business growth.

We can summarize and better highlight these practical implications as follows. First and foremost, top management teams in large hospitality firms (such as hotel chains listed on stock exchanges) and in other large corporations operating in travel and tourism (such as airline companies) need to be supported by good data analysts familiar with the latest developments of data science (Davenport and Patil, 2012). Small- and medium-sized hospitality and tourism enterprises that are not able to hire data scientists could purchase reports generated by consulting companies working on BD analytics, or find a way to collaborate and form groups to generate the critical mass needed to assemble the necessary resources for these activities. Second, more fluid communication flows should connect managers in charge of strategic decision-making with data analysts/scientists, especially if the latter ones are not “vertically specialized” in the hospitality and tourism sector. The choice of relevant data sources, suitable collection methods and data cleaning/validation and understanding techniques will stem from a fruitful dialogue between practitioners knowledgeable of the distinctive features of the business operations (i.e. the managers) and the data analysts/scientists. Third, and related to the previous point, right now – apart from a few hospitality and tourism scholars pioneering BD – there are no vertically specialized data analysts, with the exception of a handful of companies (e.g. STR Global in hospitality) that develop “analytics” in a rather traditional way based on competitive performance measures shared by company managers. There is, therefore, a huge opportunity even for those companies to involve data scientists interested in hospitality and tourism. On the other hand, there is a huge opportunity for educational institutions to revise and rejuvenate their programs in the field by dedicating resources to educate on the issues of BD and BI. Fourth, digital business models in tourism and hospitality are being developed at the speed of light through corporate entrepreneurship initiatives (see the incubators and accelerators of such companies as Travelport in the digital distribution domain); certainly those companies could capitalize even more on BD-informed products and services that might improve their competitiveness and performance.

5.4 Limitations and future research

This research is not without limitations. First, from the observed overall trends related to the scientific production related to BI (Figure 1) and BD (Figure 2), it is apparent that these two areas of research are gaining increasing scientific attention. Therefore, we expect that at the moment of writing our work, other articles are being produced and published at a very fast pace and, thus, cannot be accommodated in this review. Second, as usually done (Gomezeli, 2016; Law *et al.*, 2016), we have used the two main literature databases: Scopus and Web of Science. These, although collecting a large number of journals and scientific publications, and probably all the most important ones, still cannot be considered complete, also considering the process that takes for these repositories to index new publications. Finally, we had to simplify and reduce most of the technical details typically used in the BI and BD fields to make the article accessible to a wide social science audience. This decision, while making our work more manageable, has prevented us from going deeper into many features and technical aspects related to the fields scrutinized; such details might allow us to further enrich our research agenda.

Besides limitations, this systematic literature review has allowed us to identify several main themes and issues that might contribute to shape future research agendas for BD and

BI in hospitality and tourism. Although quite popular and strongly pushed by many, there is still a notion that BD is largely overlooked by the majority of researchers in the field. The same seems to happen for what concerns BI studies. It is difficult to fully understand the reasons for this situation, yet we interpret them as follows. First, there seems to exist a cultural gap between hospitality and tourism researchers who should be ready to embrace, accept and implement novel research methods. This needs to pass, for many, through a steep learning curve. Moreover, refinements and revisions of methodological approaches are needed, especially in terms of the combination of “old” and “new” data. Time will be needed to legitimize and sustain the aforementioned data-driven methods in the hospitality and tourism scientific community (Xiang, 2018; Shoval *et al.*, 2018). Second, and related to the previous point, scholars willing to make robust contributions to BI and BD in hospitality and tourism in the future should make sure to:

- choose suitable data sources (e.g. social media, online payment and credit card transactions, mobile phone traffic, e-Commerce transactions and booking engines, etc.);
- collect and store data on a large scale, making use of appropriate data management techniques, ranging from traditional data warehouse structures to more flexible an agile data lakes;
- clean and validate data, taking into account volatility, replicability, privacy, overlaps and redundancies;
- extract meaningful knowledge from large data volumes (by also considering suitable theoretical frameworks) in real-time, if necessary, and with an emphasis on multi- and omni-channel behaviours;
- combine different data sources for deriving complete, valid and reliable outcomes at different levels (individual, firm, business network, industry); and
- use content spontaneously generated by internet and Web users to gain additional knowledge on travellers’ beliefs, behaviours and preferences in order to overcome limitations of traditional survey-based research methods.

Third, BD and BI educational programmes and units for hospitality and tourism management have been largely missing. Therefore, academic institutions and schools should introduce, support and develop them by means of specific investments. This will be relevant to bridge a cultural gap preventing the next generation of managers to make sense of BD for BI purposes (Coleman *et al.*, 2016). Progress in tourism and hospitality research will be possible only by coaching researchers to be able to address the needs of tourism companies in a data-led economy. Fourth, the resources (hardware and software) needed to actually treat huge quantities of data are not easily available to hospitality and tourism researchers, but rather sit in computer science departments (Ekbja *et al.*, 2015). Moreover, many of the modern analysis techniques require a good knowledge of some computer programming language or database management system that are not very popular among the scholars in the tourism field. Similarly, for BI good practices call for well and rationally designed, organized and managed information systems. To overcome these issues, it is necessary for hospitality and tourism scholars to set up inter- and multi-disciplinary collaborations and research teams with an inclusive attitude towards computer and data scientists. Finally, given the issues of privacy related to retrievable data for research purposes that could be either public or confidential and proprietary, academic and research institutions have to initiate, build and leverage strong partnerships and collaborations with industry and institutional stakeholders.

References

- Alaei, A.R., Becken, S. and Stantic, B. (2017), "Sentiment analysis in tourism: capitalizing on big data", *Journal of Travel Research*, doi: [10.1177/0047287517747753](https://doi.org/10.1177/0047287517747753).
- Amadio, W.J. and Procaccino, J.D. (2016), "Competitive analysis of online reviews using exploratory text mining", *Tourism and Hospitality Management*, Vol. 22 No. 2, pp. 193-210.
- Andersson, C. (2008), "The end of theory: the data deluge makes the scientific method obsolete", *Wired Magazine*, Vol. 16 No. 7, pp. 1-3.
- Arbelaitz, O., Gurrutxaga, I., Lojo, A., Muguerza, J., Pérez, J.M. and Perona, I. (2013), "Web usage and content mining to extract knowledge for modelling the users of the bidasoa turismo website", *Expert Systems with Applications*, Vol. 40 No. 18, pp. 7478-7491.
- Ashiabor, S., Baik, H. and Trani, A. (2007), "Logit models for forecasting nationwide intercity travel demand in the United States", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2007 No. 1, pp. 1-12.
- Baggio, R. (2016), "Improving tourism statistics: merging official records with big data", in Fuchs, M., Höpken, W. and Lexhagen, M. (Eds), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund, pp. 89-92.
- Baggio, R. and Del Chiappa, G. (2014), "Real and virtual relationships in tourism digital ecosystems", *Information Technology and Tourism*, Vol. 14 No. 1, pp. 3-19.
- Barbieri, C. and Sotomayor, S. (2013), "Surf travel behavior and destination preferences: an application of the serious leisure inventory and measure", *Tourism Management*, Vol. 35, pp. 111-121.
- Bedeley, R. and Nemati, H. (2014), "Big data analytics: a key capability for competitive advantage", paper presented at the 20th Americas Conference on Information Systems (AMCIS), 7-9 August.
- Berezina, K., Bilgihan, A., Cobanoglu, C. and Okumus, F. (2016), "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews", *Journal of Hospitality Marketing & Management*, Vol. 25 No. 1 pp. 1-24.
- Boyd, D. and Crawford, K. (2012), "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication and Society*, Vol. 15 No. 5, pp. 662-679.
- Bryson, S., Kenwright, D., Cox, M., Ellsworth, D. and Haimes, R. (1999), "Visually exploring gigabyte datasets in real time", *Communications of the ACM*, Vol. 42 No. 8, pp. 83-90.
- Buhalis, D. and Foerste, M. (2015), "SoCoMo marketing for travel and tourism: empowering co-creation of value", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 151-161.
- Burstein, F. and Holsapple, C.W. (2008), *Handbook on Decision Support Systems*, Springer, New York, NY.
- Carrasco, R.A., Muñoz-Leiva, F. and Hornos, M.J. (2013), "A multidimensional data model using the fuzzy model based on the semantic translation", *Information Systems Frontiers*, Vol. 15 No. 3, pp. 351-370.
- Carter, P.L. (2016), "Where are the enslaved?: tripadvisor and the narrative landscapes of Southern plantation museums", *Journal of Heritage Tourism*, Vol. 11 No. 3, pp. 235-249.
- Chang, R.M., Kauffman, R.J. and Kwon, Y. (2014), "Understanding the paradigm shift to computational social science in the presence of big data", *Decision Support Systems*, Vol. 63, pp. 67-80.
- Chen, L.F. and Tsai, C.T. (2016), "Data mining framework based on rough set theory to improve location selection decisions: a case study of a restaurant chain", *Tourism Management*, Vol. 53, pp. 197-206.
- Chen, M., Mao, S. and Liu, Y. (2014), "Big data: a survey", *Mobile Networks and Applications*, Vol. 19 No. 2, pp. 171-209.

-
- Chiang, W.-Y. (2015), "Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan", *Transportation Letters*, Vol. 6 No. 2, pp. 89-97.
- Chugh, R. and Grandhi, S. (2013), "Why business intelligence? Significance of business intelligence tools and integrating BI governance with corporate governance", *International Journal of E-Entrepreneurship and Innovation*, Vol. 4 No. 2, pp. 1-14.
- Coker, F. (2014), *Pulse: Understanding the Vital Signs of Your Business*, Ambient Light Publishing, Bellevue, WA.
- Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X. and Seabra Reis, M. (2016), "How can SMEs benefit from big data? Challenges and path forward", *Quality and Reliability Engineering International*, Vol. 32 No. 6, pp. 2151-2164.
- Correia, A., Kozak, M. and Ferradeira, J. (2013), "From tourist motivations to tourist satisfaction. International", *International Journal of Culture, Tourism and Hospitality Research*, Vol. 7 No. 4, pp. 411-424.
- d'Amore, M., Baggio, R. and Valdani, E. (2015), "A practical approach to big data in tourism: a low cost raspberry pi cluster", In Tussyadiah, I. and Inversini, A. (Eds), *Information and Communication Technologies in Tourism 2015 (Proceedings of the International Conference in Lugano, Switzerland, February 3-6)*, Springer, Berlin, pp. 169-181.
- D'Aveni, R.A. (1994), *Hyper Competition: Managing the Dynamics of Strategic Maneuvering*, Free Press, New York, NY.
- D'Aveni, R.A., Dagnino, G.B. and Smith, K.G. (2010), "The age of temporary advantage", *Strategic Management Journal*, Vol. 31 No. 13, pp. 1371-1385.
- Davenport, T.H. and Patil, D.J. (2012), "Data scientist: the sexiest job of the 21st century", *Harvard Business Review*, Vol. 90 No. 10, pp. 70-76.
- Dedić, N. and Stanier, C. (2016), "Measuring the success of changes to existing business intelligence solutions to improve business intelligence reporting", *Lecture Notes in Business Information Processing*, Springer International Publishing, Vol. 268, pp. 225-236.
- Dergiades, T., Mavragani, E. and Pan, B. (2018), "Google trends and tourists' arrivals: emerging biases and proposed corrections", *Tourism Management*, Vol. 66, pp. 108-120.
- Devens, R.M. and Miller, R. (2013), *Encyclopaedia of Commercial and Business Anecdotes*, Hard Press Publishing, New York, NY, Vol. 1.
- Dolnicar, S. and Ring, A. (2014), "Tourism marketing research: past, present and future", *Annals of Tourism Research*, Vol. 47, pp. 31-47.
- Dursun, A. and Caber, M. (2016), "Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis", *Tourism Management Perspectives*, Vol. 18, pp. 153-160.
- Ekbia, H., Mattioli, M., Arave, G., Ghazinejad, A., Bowman, T., Suri, V., Tsou, A., Weingart, S. and Sugimoto, C.R. (2015), "Big data, bigger dilemmas: a critical review", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 8, pp. 1523-1545.
- Erl, T., Khattak, W. and Buhler, P. (2015), *Big Data Fundamentals – Concepts, Drivers and Techniques*, Prentice Hall, New York, NY.
- Fan, J., Han, F. and Liu, H. (2014), "Challenges of big data analysis", *National Science Review*, Vol. 1 No. 2, pp. 293-314.
- Fan, S., Lau, R.Y. and Zhao, J.L. (2015), "Demystifying big data analytics for business intelligence through the lens of marketing mix", *Big Data Research*, Vol. 2 No. 1, pp. 28-32.
- Franks, B. (2012), *Teaming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, John Wiley and Sons, Hoboken, NJ.
- Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W. and Lexhagen, M. (2013), "A knowledge destination framework for tourism sustainability: a business intelligence application from Sweden", *Tourism: An Interdisciplinary Journal*, Vol. 61 No. 2, pp. 121-148.

- Fuchs, M., Höpken, W. and Lexhagen, M. (2016), "Dynamic need fulfilment in a collaborative destination environment", in Fuchs, M., Höpken, W. and Lexhagen, M. (Eds.), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund, pp. 97-100.
- Fuchs, M., Höpken, W. and Lexhagen, M. (2014), "Big data analytics for knowledge generation in tourism destinations – a case from Sweden", *Journal of Destination Marketing and Management*, Vol. 3 No. 4, pp. 198-209.
- Gandomi, A. and Haider, M. (2015), "Beyond the hype: big data concepts, methods, and analytics", *International Journal of Information Management*, Vol. 35 No. 2, pp. 137-144.
- García-Palomares, J.C., Gutiérrez, J. and Mínguez, C. (2015), "Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS", *Applied Geography*, Vol. 63, pp. 408-417.
- García-Pablos, A., Cuadros, M. and Linaza, M.T. (2016), "Automatic analysis of textual hotel reviews", *Information Technology and Tourism*, Vol. 16 No. 1, pp. 45-69.
- Gerard, G., Osinga, E.C., Lavie, D. and Scott, B.A. (2016), "Big data and data science methods for management research", *Academy of Management Journal*, Vol. 59 No. 5, pp. 1493-1507.
- Gomezelj, D.O. (2016), "A systematic review of research on innovation in hospitality and tourism", *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 3, pp. 516-558.
- Gong, L., Liu, X., Wu, L. and Liu, Y. (2016), "Inferring trip purposes and uncovering travel patterns from taxi trajectory data", *Cartography and Geographic Information Science*, Vol. 43 No. 2, pp. 103-114.
- Granville, V. (2013), "The curse of big data", available at: www.analyticbridge.com/profiles/blogs/the-curse-of-big-data (accessed June 2015).
- Gretzel, U., Sigala, M., Xiang, Z. and Koo, C. (2015), "Smart tourism: foundations and developments", *Electronic Markets*, Vol. 25 No. 3, pp. 179-188.
- Gunter, U. and Önder, I. (2016), "Forecasting city arrivals with Google analytics", *Annals of Tourism Research*, Vol. 61, pp. 199-212.
- Han, B.-C. (2015), *Psychopolitik – Neoliberalismus Und Die Neuen Machttechniken*, Fischer, Frankfurt a.M.
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P. and Ratti, C. (2014), "Geo-located twitter as proxy for global mobility patterns", *Cartography and Geographic Information Science*, Vol. 41 No. 3, pp. 260-271.
- He, J., Liu, H. and Xiong, H. (2016), "SocoTraveler: travel-package recommendations leveraging social influence of different relationship types", *Information and Management*, Vol. 53 No. 8, pp. 934-950.
- Heerschap, N., Ortega, S., Priem, A. and Offermans, M. (2014), "Innovation of tourism statistics through the use of new big data sources", Paper presented at the 12th Global Forum on Tourism Statistics, 15-16 May 2014, Prague, Czech Republic, available at: www.tsf2014prague.cz/assets/downloads/Paper%201.2_Nicolaes%20Heerschap_NL.pdf (accessed July 2014).
- Hegel, G.W.F. (1830), *Enzyklopädie Der Philosophischen Wissenschaften Im Grundrisse – Die Wissenschaft Der Logik*, Frankfurt a. M.
- Hermans, E. and Liu, F. (2016), "Using and combining promising data types in transportation models", in Fuchs, M., Höpken, W. and Lexhagen, M. (Eds.), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund, pp. 101-103.
- Holland, C.P., Jacobs, J.A. and 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 Tourism*, Vol. 16 No. 1, pp. 127-148.

-
- Höpken, W., Ph, D., Höll, G., Kuppe, J., Schorpp, D., Licones, R. and Fuchs, M. (2012), "Digitalizing loyalty cards in tourism", in Fuchs, M., Ricci, F. and Cantoni, L. (Eds), *Information and Comm. Tech. in Tourism*, Springer, New York, NY, pp. 272-283.
- Höpken, W., Ernesti, D., Fuchs, M., Kronenberg, K. and Lexhagen, M. (2017), "Big data as input for predicting tourist arrivals", In Schegg, R. and Stangl, B. (Eds), *Information and Communication Technologies in Tourism 2017*, Springer, New York, NY, 187-200.
- Höpken, W., Fuchs, M., Keil, D. and Lexhagen, M. (2015), "Business intelligence for cross-process knowledge extraction at tourism destinations", *Information Technology and Tourism*, Vol. 15 No. 2, pp. 101-130.
- Hsieh, K. (2009), "Applying an expert system into constructing customer's value expansion and prediction model based on AI techniques in leisure industry", *Expert Systems with Applications*, Vol. 36 No. 2, PART 2, pp. 2864-2872.
- Hsieh, K. (2011), "Employing a recommendation expert system based on mental accounting and artificial neural networks into mining business intelligence for study abroad' s P/S recommendations", *Expert Systems with Applications*, Vol. 38 No. 12, pp. 14376-14381.
- IBM (2012), "Analytics. The real world use of Big Data – How innovative enterprises extract value from uncertain data IBM Institute for Business Value", available at: www.informationweek.com/pdf_whitepapers/approved/1372892704_analytics_the_real_world_use_of_big_data.pdf (accessed June 27th).
- Ip, C., Law, R. and Lee, H.A. (2011), "A review of website evaluation studies in the tourism and hospitality fields from 1996 to 2009", *International Journal of Tourism Research*, Vol. 13 No. 3, pp. 234-265.
- Jackson, S. (2016), "Prediction, explanation and big(ger) data: a middle way to measuring and modelling the perceived success of a volunteer tourism sustainability campaign based on 'nudging'", *Current Issues in Tourism*, Vol. 19 No. 7, pp. 643-658.
- Jannach, D. (2016), "Big data for travel and tourism recommender systems: a position statement", in Fuchs, M., Höpken, W. and Lexhagen, M. (Eds), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund, pp. 109-110.
- Kasahara, H., Mori, M., Mukunoki, M. and Minoh, M. (2015), "Transportation mode annotation of tourist GPS trajectories under environmental constraints", in Inversini, A. and Schegg, R. (Eds), *Information and Communication Technologies in Tourism 2015*, Springer, Heidelberg, pp. 523-535.
- Kimball, R. and Ross, M. (2016), *The Kimball Group Reader – Relentlessly Practical Tools for Data Warehousing and Business Intelligence*, (2nd ed.), John Wiley and Sons, Indianapolis.
- Kimball, R., Ross, M., Thronthwaite, W., Mundy, J. and Becker, B. (2008), *The Data Warehouse Lifecycle Toolkit*, (2nd ed.), John Wiley and Sons, Indianapolis.
- Kirilenko, A.P., Stepchenkova, S.O., Kim, H. and Li, R.X. (2017), "Automated sentiment analysis in tourism: Comparison of approaches", *Journal of Travel Research*, doi: [10.1177/0047287517729757](https://doi.org/10.1177/0047287517729757).
- Kisilevich, S., Keim, D. and Rokach, L. (2013), "A GIS-based decision support system for hotel room rate estimation and temporal price prediction: the hotel brokers' context", *Decision Support Systems*, Vol. 54 No. 2, pp. 1119-1133.
- Kitchin, R. and Lauriault, T.P. (2015), "Small data in the era of big data", *GeoJournal*, Vol. 80 No. 4, pp. 463-475.
- Klein, S. and Jacobs, J.A. (2016), "Position statement on big data and business intelligence in tourism", in Fuchs, M., Höpken, W. and Lexhagen, M. (Eds), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund, pp. 105-108.
- Knudsen, C. (2003), "Pluralism, scientific progress, and the structure of organization theory", in Tsoukas, H. and Knudsen, C. (Eds), *The Oxford Handbook of Organization Theory: Meta-Theoretical Perspectives*, Oxford University Press, Oxford, pp. 262-286.

- Kolas, N., Höpken, W., Fuchs, M. and Lexhagen, M. (2015), "Information gathering by ubiquitous services for CRM in tourism destinations", in Tussyadiah, I. and Inversini, A. (Eds), *Information and Communication Technologies in Tourism*, Springer, New York, NY, pp. 73-85.
- Kong, H. and Song, E. (2016), "A study on customer feedback of tourism service using social big data", *International Journal of u- and e-Service, Science and Technology*, Vol. 9 No. 7, pp. 49-54.
- Köseoglu, M.A., Ross, G. and Okumus, F. (2016), "Competitive intelligence practices in hotels", *International Journal of Hospitality Management*, Vol. 53, pp. 161-172.
- Kwok, L. and Yu, B. (2016), "Taxonomy of facebook messages in business-to-consumer communications: What really works?", *Tourism and Hospitality Research*, Vol. 16 No. 4, pp. 311-328.
- Lam, C. and McKercher, B. (2013), "The tourism data gap: the utility of official tourism information for the hospitality and tourism industry", *Tourism Management Perspectives*, Vol. 6, pp. 82-94.
- Laney, D. (2001), "3D data management: controlling data volume, velocity and variety", *Meta Group Research Note*, Vol. 6 No. 70, p. 1, available at: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Larose, D.T. (2005), *Discovering Knowledge in Data – an Introduction to Data Mining*, John Wiley and Sons, Hoboken, NJ.
- Law, R., Rong, J., Vu, H., Li, G. and Lee, A. (2011), "Identifying changes and trends in Hong Kong outbound tourism", *Tourism Management*, Vol. 32 No. 5, pp. 1106-1114.
- Law, R., Sun, S., Ka, D., Fong, C., Hoc, L., Fong, N. and Fu, H. (2016), "A systematic review of china's outbound tourism research", *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 12, pp. 2654-2674.
- Lazer, D.M., Kennedy, R., King, G. and Vespignani, A. (2014), "The parable of google flu: traps in big data analysis", *Science (New York, N.Y.)*, Vol. 343 No. 6176, pp. 1203-1205.
- Leung, R., Rong, J., Li, G. and Law, R. (2013), "Personality differences and hotel web design study using targeted positive and negative association rule mining", *Journal of Hospitality Marketing and Management*, Vol. 22 No. 7, pp. 701-727.
- Li, G., Law, R., Vu, H.Q., Rong, J. and Zhao, X. (2015), "Identifying emerging hotel preferences using emerging pattern mining technique", *Tourism Management*, Vol. 46, pp. 311-321.
- Liebowitz, J. (Ed.) (2013), *Big Data and Business Analytics*, CRC Press, Boca Raton, FL.
- Liu, F., Janssens, D., Wets, G. and Cools, M. (2013), "Annotating mobile phone location data with activity purposes using machine learning algorithms", *Expert Systems with Applications*, Vol. 40 No. 8, pp. 3299-3311.
- Lu, W. and Stepchenkova, S. (2015), "User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software", *Journal of Hospitality Marketing and Management*, Vol. 24 No. 2, pp. 119-154.
- Lu, Y. and Zhang, L. (2015), "Imputing trip purposes for long-distance travel", *Transportation*, Vol. 42 No. 4, pp. 581-595.
- Luhn, H.P. (1958), "A business intelligence system", *IBM Journal of Research and Development*, Vol. 2 No. 4, pp. 314-319.
- Lycett, M. (2013), "Datafication': making sense of (big) data in a complex world", *European Journal of Information Systems*, Vol. 22 No. 4, pp. 381-386.
- McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D. (2012), "Big data. The management revolution", *Harvard Business Review*, Vol. 90 No. 10, pp. 61-67.
- McFarland, D.A. and McFarland, H.R. (2015), "Big data and the danger of being precisely inaccurate", *Big Data and Society*, Vol. 2 No. 2, pp. 1-4.
- Mariani, M. and Baggio, R. (2012), "Special issue: managing tourism in a changing world: issues and cases", *Anatolia: An International Journal of Tourism and Hospitality Research*, Vol. 23 No. 1 pp. 1-3.

- Mariani, M.M. and Borghi, M. (2018), "Effects of the booking.com rating system: binging hotel class into the picture", *Tourism Management*, Vol. 66, pp. 47-52.
- Mariani, M.M., Borghi, M. and Gretzel, U. (2019), "Online reviews: differences by submission device", *Tourism Management*, Vol. 70, pp. 295-298.
- Mariani, M.M., Buhalis, D., Czakon, W. and Vitouladiti, O. (2016), *Tourism Management, Marketing, and Development. Performance, Strategies, and Sustainability*, Palgrave-Macmillan, New York, NY.
- Mariani, M.M., Di Felice, M. and Mura, M. (2016), "Facebook as a destination marketing tool: evidence from italian regional destination management organizations", *Tourism Management*, Vol. 54, pp. 321-343.
- Mariani, M.M., Baggio, R., Buhalis, D. and Longhi, C. (2014), in Mariani, M.M., Baggio, R., Buhalis, D. and Longhi, C. (Eds), *Tourism Management, Marketing and Development: the Importance of Networks and ICTs*, New York: Palgrave, pp. 1-11, doi: [10.1057/9781137354358](https://doi.org/10.1057/9781137354358).
- Marine-Roig, E. and Anton Clavé, S.A. (2015), "Tourism analytics with massive user-generated content: a case study of barcelona", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 162-172.
- Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, New York, NY.
- Menner, T., Höpken, W., Fuchs, M. and Lexhagen, M. (2016), "Topic detection – identifying relevant topics, within touristic UGC", in Inversini, A. and Schegg, R. (Eds), *Information and Communication Technologies in Tourism 2015*, Springer, Heidelberg, pp. 411-423.
- Mi, C., Shan, X., Qiang, Y., Stephanie, Y. and Chen, Y. (2014), "A new method for evaluating tour online review based on grey 2-tuple linguistic", *Kybernetes*, Vol. 43 Nos 3/4, pp. 601-613.
- Mitchell, T.M. (1997), *Machine Learning*, McGraw Hill, New York, NY.
- Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. and Vespignani, A. (2013), "The twitter of babel: mapping world languages through microblogging platforms", *PLoS ONE*, Vol. 8 No. 4, doi: [10.1371/journal.pone.0061981](https://doi.org/10.1371/journal.pone.0061981).
- Noguchi, K., Sato, Y. and Shiohara, H. (2016), "Advanced, high-performance big data technology and trial validation", *NTT Technical Review*, Vol. 14 No. 2, pp. 1-7.
- Önder, I., Koerbitz, W. and Hubmann-Haidvogel, A. (2014), "Tracing tourists by their digital footprints the case of Austria", *Journal of Travel Research*, Vol. 55 No. 5, pp. 566-573. doi: [10.1177/0047287514563985](https://doi.org/10.1177/0047287514563985).
- Orellana, D., Bregt, A.K., Ligtenberg, A. and Wachowicz, M. (2012), "Exploring visitor movement patterns in natural recreational areas", *Tourism Management*, Vol. 33 No. 3, pp. 672-682.
- Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C. and González, M.C. (2015), "Urban magnetism through the lens of geo-tagged photography", *EPJ Data Science*, Vol. 4 No. 1, pp. 1-17.
- Park, S.B., Ok, C.M. and Chae, B.K. (2016), "Using twitter data for cruise tourism marketing and research", *Journal of Travel and Tourism Marketing*, Vol. 33 No. 6, pp. 885-898. doi: [10.1080/10548408.2015.11071688](https://doi.org/10.1080/10548408.2015.11071688).
- Pope, S., Garrow, L., Guin, A., Leonard, J., Bankston, L. and Campbell, P. (2009), "Conceptual framework for collecting online airline pricing data: challenges, opportunities, and preliminary results", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2106 No. 1, pp. 30-37.
- Power, D.J. (2007), "A brief history of decision support systems", version 4.0, available at: DSS-Resources.com (accessed 6 June 2017).
- Power, D.J. (2014), "Using 'big data' for analytics and decision support", *Journal of Decision Systems*, Vol. 23 No. 2, pp. 222-228.
- Prayag, G., Hosany, S. and Odeh, K. (2013), "The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions", *Journal of Destination Marketing and Management*, Vol. 2 No. 2, pp. 118-127.

- Pyle, D. (1999), *Data Preparation for Data Mining*, Morgan Kaufmann, Oxford.
- Qiao, X., Zhang, L., Li, N. and Zhu, W. (2014), "Constructing a data warehouse based decision support platform for China tourism industry", in Xiang, Z. and Tussyadiah, I. (Eds), *Information and Communication Technologies in Tourism 2014*, Springer, Heidelberg, pp. 883-893.
- Raun, J., Ahas, R. and Tiru, M. (2016), "Measuring tourism destinations using mobile tracking data", *Tourism Management*, Vol. 57, pp. 202-212.
- Ritchie, R.J. and Ritchie, J.B. (2002), "A framework for an industry supported destination marketing information system", *Tourism Management*, Vol. 23 No. 5, pp. 439-454.
- RocaSalvatella (2014), *Big Data and Tourism: New Indicators for Tourism Management*, Telefónica I+D and RocaSalvatella, Barcelona, available at: www.rocasalvatella.com/sites/default/files/big_data_y_turismo-eng-interactivo.pdf (accessed September 2015).
- Rossetti, M., Stella, F. and Zanker, M. (2016), "Analyzing user reviews in tourism with topic models", *Information Technology and Tourism*, Vol. 16 No. 1, pp. 5-21.
- Rud, O. (2009), *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy*, Wiley and Sons, N.J.
- Russell, M.A. (2013), *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*, O'Reilly Media Inc, Sebastopol, CA.
- Sánchez-Franco, M.J., Navarro-García, A. and Rondán-Cataluña, F.J. (2016), "Online customer service reviews in urban hotels: a data mining approach", *Psychology and Marketing*, Vol. 33 No. 12, pp. 1174-1186.
- Sauter, V. (2011), *Decision Support Systems for Business Intelligence*, (2nd edition), John Wiley and Sons, Hoboken, NJ.
- Schmunk, S., Höpken, W., Fuchs, M. and Lexhagen, M. (2014), "Sentiment analysis – implementation and evaluation of methods for sentiment analysis with rapid-Miner®", in Xiang, Z. and Tussyadiah, I. (Eds), *Information and Communication Technologies in Tourism 2014*, Springer, Heidelberg, pp. 253-265.
- Shi, M., Zhu, W., Yang, H. and Li, C. (2016), "Applying semantic web and big data tech-niques to construct a balance model referring to stakeholders of tourism intangible cultural heritage", *International Journal of Computer Applications in Technology*, Vol. 54 No. 3, pp. 192-200.
- Shoval, N. and Isaacson, M. (2010), *Tourist Mobility and Advanced Tracking Technologies*, Routledge, New York, NY.
- Shoval, N., Schvimer, Y. and Tamir, M. (2018), "Real-time measurement of tourists' objective and subjective emotions in time and space", *Journal of Travel Research*, Vol. 57 No. 1, pp. 3-16.
- Sinisalo, J., Salo, J., Karjalainen, H. and Leppäniemi, M. (2007), "Mobile customer relationship management: underlying issues and challenges", *Business Process Management Journal*, Vol. 13 No. 6, pp. 771-787.
- Snaveley, N., Seitz, S.M. and Szeliski, R. (2008), "Modelling the world from internet photo collections", *International Journal of Computer Vision*, Vol. 80 No. 2, pp. 189-210.
- Solnet, D., Boztug, Y. and Dolnicar, S. (2016), "An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue", *International Journal of Hospitality Management*, Vol. 56 No. 3, pp. 119-125.
- Su, S., Wan, C., Hu, Y. and Cai, Z. (2016), "Characterizing geographical preferences of international tourists and the local influential factors in china using geo-tagged photos on social media", *Applied Geography*, Vol. 73, pp. 26-37.
- Sun, P., Cárdenas, D.A. and Harill, R. (2016), "Chinese customers' evaluation of travel website quality: a decision-Tree Analysis", *Journal of Hospitality Marketing and Management*, Vol. 25 No. 4, pp. 476-497.
- Sun, Y., Song, H., Jara, A.J. and Bie, R. (2016), "Internet of things and big data analytics for smart and connected communities", *IEEE Access*, Vol. 4, pp. 766-773.

-
- Supak, S., Brothers, G., Bohnenstiehl, D. and Devine, H. (2015), "Geospatial analytics for federally managed tourism destinations and their demand markets", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 173-186.
- Tang, J. and Li, J. (2016), "Spatial network of urban tourist flow in xi'an based on microblog big data", *Journal of China Tourism Research*, Vol. 12 No. 1, pp. 5-23.
- Tranfeld, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", *British Journal of Management*, Vol. 14 No. 3, pp. 207-222.
- Tseng, S. and Won, Y. (2016), "Integrating multiple recommendation schemes for designing sales force support system: a travel agency example", *International Journal of Electronic Business*, Vol. 13 No. 1, pp. 1-37.
- Verhoef, P.C., Kooge, E. and Walk, N. (2016), *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions*, Routledge, London.
- Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I. and 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*, Vol. 44, pp. 67-81.
- Vieira, E.S. and Gomes, J.A.N.F. (2009), "A comparison of scopus and web of science for a typical university", *Scientometrics*, Vol. 81 No. 2, pp. 587-600.
- Vogt, C. (2011), "Customer relationship management in tourism", *Journal of Travel Research*, Vol. 50 No. 4, pp. 356-364.
- Yang, Y., Liao, L., Yan, L., Hu, X-C, Huang, H-B and Xiao, S. (2016), "The big data analysis of land use evolution and its ecological security responses in Silver Beach of China by the clustering of spatial patterns", *Cluster Computing*, Vol. 19 No. 4, pp. 1907-1924.
- Wang, D., Park, S. and Fesenmaier, D.R. (2012), "The role of smartphones in mediating the touristic experience", *Journal of Travel Research*, Vol. 51 No. 4, pp. 371-387.
- Wang, X.L., Yoonjoung Heo, C., Schwartz, Z., Legoh  rel, P. and Specklin, F. (2015), "Revenue management: progress, challenges, and research prospects", *Journal of Travel and Tourism Marketing*, Vol. 32 No. 7, pp. 797-811.
- White, T. (2015), "Hadoop: the definitive guide", 4th Edition - *Storage and Analysis at Internet Scale*, O'Reilly Media, Sebastopol, CA.
- Williams, S. (2016), *Business Intelligence Strategy and Big Data Analytics*, Morgan Kaufmann, Cambridge (MA).
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Cambridge (MA).
- Wood, S.A., Guerry, A.D., Silver, J.M. and Lacayo, M. (2013), "Using social media to quantify nature-based tourism and recreation", *Scientific Reports*, Vol. 3 No. 1, doi: [10.1038/srep02976](https://doi.org/10.1038/srep02976).
- Wu, E.H.C., Law, R. and Jiang, B. (2010), "Data mining for hotel occupancy rate: an independent component analysis approach", *Journal of Travel and Tourism Marketing*, Vol. 27 No. 4, pp. 426-438.
- Xiang, Z. (2016), "Analytics for tourism management: Needs and directions for research", in Fuchs, M., H  pken, W. and Lexhagen, M. (Eds), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University,   stersund, pp. 127-131.
- Xiang, Z. (2018), "From digitization to the age of acceleration: on information technology and tourism", *Tourism Management Perspectives*, Vol. 25, pp. 147-150, doi.org/10.1016/j.tmp.2017.11.023.
- Xiang, Z., Schwartz, Z., Gerdes, J.H. and Uysal, M. (2015), "What can big data and text analytics tell us about hotel guest experience and satisfaction?", *International Journal of Hospitality Management*, Vol. 44, pp. 120-130.
- Yang, Y., Pan, B. and Song, H. (2014), "Predicting hotel demand using destination marketing organization's web traffic data", *Journal of Travel Research*, Vol. 53 No. 4, pp. 433-447.

- Yuan, Y.L. and Ho, C.I. (2015), "Rethinking the destination marketing organization management in the big data era", in *Proceedings of the ASE Big Data and Social Informatics 2015, ACM*, p. 60.
- Zanker, M., Jessenitschnig, M. and Fuchs, M. (2010), "Automated semantic annotation of tourism resources based on geo-Spatial Data", *Information Technology and Tourism*, Vol. 11 No. 4, pp. 341-354.
- Zhang, C. and Huang, Z. (2015), "Mining tourist motive for marketing development via twice-learning", *Applied Artificial Intelligence*, Vol. 29 No. 2, pp. 119-132.
- Zhu, Z., Shou, L. and Chen, K. (2016), "Get into the spirit of a location by mining user-generated travelogues", *Neurocomputing*, Vol. 204, pp. 61-69.

Further reading

- Mariani, M.M., Mura, M. and Di Felice, M. (2018), "The determinants of Facebook social engagement for national tourism organizations' Facebook pages: a quantitative approach", *Journal of Destination Marketing & Management*, Vol. 8, pp. 312-325.
- Savannah, G.A., Berezina, K., Bilgihan, A., Cobanoglu, C. and Okumus, F. (2016), "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews", *Journal of Hospitality Marketing and Management*, Vol. 25 No. 1, pp. 1-24.

Corresponding author

Marcello Mariani can be contacted at: m.mariani@henley.ac.uk