# Big Data and analytics in tourism and hospitality: a perspective article

Marcello Mariani

**Abstract**

**Purpose** – *This study aims to discuss the evolution of Big Data (BD) and Analytics in the tourism and hospitality field. It analyses the important role that BD has played so far in tourism and hospitality research and delineates how it might evolve in the future.*

**Design/methodology/approach** – *In line with the Platinum Jubilee Special Issue of* Tourism Review*, this work consists of a critical and conceptual analysis including a mini literature review of recent work in the area at the intersection of BD and tourism and hospitality research.*

**Findings** – *Findings suggest that tourism and hospitality scholars are increasingly aware of and adopting BD approaches to retrieve, collect, analyse, report and visualise their data. However, a number of avenues for improvement in the use and interpretation of BD and BD analytics as both sets of methods and technology need to be developed. Moreover, BD analytics promise to enhance a number of digital technologies in tourism and hospitality such as AI and IoT that heavily rely on data. As such, the authors envision that a new digital entrepreneurship field might be shaped within the tourism and hospitality literature. Research pathways for future inquiry at the intersection of BD and tourism and hospitality are outlined.*

**Originality/value** – *While thinking retrospectively about research revolving around BD and its role in the tourism and hospitality research field so far, this study also addresses the challenges pertaining to how BD research will be conducted in the next seven decades within tourism and hospitality.*

**Keywords** *Big Data, Tourism, Hospitality, Big Data analytics, Platinum jubilee*

**Paper type** *Research paper*

Marcello Mariani is based at the Department of Leadership, Organisations Behavior, University of Reading, Henley-on-Thames, UK.

## 1. Introduction

Every year human beings and machines increasingly generate vast amounts of data (Hilbert and López, 2011). Today, more than the 2.5 Exabytes of data are produced *daily* (IBM, 2016). The advent and consolidation of the Internet, the proliferation of smartphones and Internet of Things (IoT) devices and sensors is contributing to reinforce this growing trend. As such, we live in a data world and businesses in a data economy where data is actually the new oil propelling economic growth (The Economist, 2017). The tourism and hospitality (T&H) sectors are not an exception and tourism destinations, firms and consumers increasingly create and deploy large volumes of data to improve their decision making processes and co-create value (Mariani *et al.*, 2014; Mariani *et al.*, 2018b; Pauleen and Wang, 2017; Pigni *et al.*, 2016). For instance, tourists' location data tracked by means of GPS sensors embedded in smartphones can provide destination marketers with meaningful insights about the most popular routes and attractions in a destination; when matched with social media data, location data allow marketing firms to create location-specific offers based on the preferences and behaviours that the consumer has revealed on social media in the past, thus allowing businesses to interact in real time with potential customers based on the opinions and actions that they share and on the history of their preferences and tastes. Furthermore, online travel agencies can leverage on large volumes of structured and unstructured data stemming from online travel reviews to better

Received 22 June 2019
Revised 10 July 2019
Accepted 12 July 2019

understand tourists' online behaviours by means of digital data analytics (Mariani *et al.*, 2019c). This work looks retrospectively at the way this overabundance of travel and tourism related data, namely, Big Data (BD), has been made sense of, and examined in T&H literature and illustrates future research directions.

## 2. Past perspective 75 years of developments 1946-2020

The "Big Data" (BD) concept appeared in the late Nineties in computer science literature in relation to scientific visualisation (Cox and Ellsworth, 1997). However, its first definition was given in 2001 by Doug Laney who identified three major characteristics of BD as the 3Vs: Volume (size of data, in the order of Zettabytes), Velocity (rapidity of data generation, modification and transfer), and Variety (data can assume different formats/structures). Later the definitional model was perfected by introducing the Vs of Value (the process of extracting valuable knowledge from data by means of BD analytics) and Veracity (the governance of data in relation to their reliability) thus leading to the formulation of a 5Vs framework (Bello-Orgaz *et al.*, 2016).

As observed in a recent systematic literature review (Mariani *et al.*, 2018a), research on the role of BD in T&H is still rather fragmented and "relegated to isolated research questions" (Mariani *et al.*, 2018a: 3516). Based on the aforementioned literature review matched with a further review of articles published in the years 2017-2019, we draw some reflections on extant research. First, scholars seem increasingly interested in adopting large volumes of data to better understand complex T&H phenomena as witnessed by an increase in the number of articles explicitly adopting BD and BD analytics (Liu *et al.*, 2018; Mariani and Borghi, 2018; Talón-Ballestero *et al.*, 2018). However this body of research is a marginal share of the studies produced in the T&H field. Second, while articles have adopted different sources of data, the majority of BD articles in tourism relies on one specific source which is UGC data (e.g. social media post and online travel reviews), followed by location data (e.g. GPS data), and travel information search data. This is certainly not in line with the notion that BD studies should employ and triangulate a variety of sources and formats of data (Sun *et al.*, 2019). Third, in terms of techniques deployed, while traditional regression analyses are still dominant, increasingly text analytics are deployed, as well as machine learning techniques. However, artificial intelligence or Bayesian classification methods are extremely rare. Fourth, as far as research streams are concerned, it seems that eWOM research is the most reliant on the use of BD to dig in depth about online travellers' preferences, behaviours, satisfaction and mobility (Mariani and Borghi, 2018) and increasingly about firms' efficiency (Mariani and Visani, 2019). Last, it seems that analytics (especially text analytics) are becoming increasingly relevant in a number of studies (Xiang *et al.*, 2017; Zhao *et al.*, 2019).

## 3. Future perspective 75 years 2020-2095

It is envisioned that the research field of T&H will increasingly embrace BD and BD analytics at different levels, speed and for different purposes. In particular, BD will increasingly contribute to:

- frame novel research questions and hypotheses if combined with an underpinning conceptual framework;
- enrich research designs and methods;
- improve the generalizability of research findings across different institutional, economic, social and geographical contexts;
- generate relevant managerial insights and business intelligence by means of (digital) data analytics in real time; and
- advance BD technological applications in the verticals of T&H.

However, to make the aforementioned contributions and seise the related opportunities, there are a few challenges that need to be addressed. First, T&H scholars will still need to rely on an overarching conceptual framework identifying critical business problems in T&H (Fuchs *et al.*, 2016) despite early attempts to develop it (*Stylos and Zwiegelaar*, 2019). As such, BD will likely bring to a situation whereby theory and data analytics capabilities (Davenport, 2017) will need to further consolidate and corroborate an emerging fertile dialogue. However, to address the aforementioned challenge, scholars should increasingly work in cross-disciplinary research teams including data and computer scientists (Mariani *et al.*, 2018a). Second, BD offer the opportunity to enrich research designs and methods as they might be progressively embedded in mixed methods (Creswell, 2014) such as explanatory/exploratory sequential designs whereby big and small data might be juxtaposed, combined and triangulated. This approach to data might call for the introduction of a new data lifecycle paradigm (Blazquez and Domenech, 2018). In terms of methods, unsupervised machine learning will probably gain momentum as well as Artificial Intelligence (AI). Third, while the retrieval of large volumes of varied data in real time might generate an opportunity to overcome generalisation issues and work with representative samples (Gerard *et al.*, 2016), the challenge would be to ensure comparability: for instance scholars interested in e-WOM will increasingly resort to cross-platform cross-country studies whereby online customer behaviours will be compared (Mariani *et al.*, 2019b) also controlling for mobile devices (Mariani *et al.*, 2019a). Fourth, increasingly applied researchers might have access to (digital) data streams (Pigni *et al.*, 2016) in real time. This might contribute to reduce the theory-practice divide in tourism management research through the generation of data-driven knowledge about markets and prospective customers that could be made actionable (in real time) by policy makers, destination managers (Deng and Li, 2018), firms' executives, and entrepreneurs to create real time T&H services (*Buhalis and Sinarta*, 2019) . However, this would imply an increasing integration of BD and business intelligence applications (Fan *et al.*, 2015). Fifth, BD analytics promise to enhance a number of digital technologies such as AI and the IoT that heavily rely on data. Indeed, beyond chatbots that are very rudimentary forms of AI, more advanced AI forms and applications increasingly need to be fed by BD to trigger learning processes that machines can use to learn from past and present customer behaviours to anticipate and identify future tourists' needs and engage with travellers. As such we envision that a new digital entrepreneurship field (Nambisan, 2017) might be shaped within the T&H literature by focussing on the disruptions brought about by digital entrepreneurs in T&H services industries (Buhalis *et al.*, 2019). This is particularly relevant as BD analytics is one of the digital technologies underpinning the Industry 4.0 (Rüßmann *et al.*, 2015). The intertwined development of AI applied to data mining and predictive analytics might translate into developments of effective digital business models to support product, process and business model innovation.

## 4. Conclusions

In conclusion, there is an increasing scholarly awareness in the T&H scientific domain that the complexity of today's phenomena needs to be addressed be leveraging BD and BD analytics. It is envisioned that in the upcoming years BD will be increasingly adopted to enrich and corroborate research designs and generate findings that could be generalizable across different geographical and social settings. Data science skills and competences should be therefore embedded into cross-disciplinary research teams. The intertwined development of BD with AI and IoT technological paradigms will call for an even more integrated approach to the study of digital transformation and entrepreneurship within T&H services.

## References


Bello-Orgaz, G., Jung, J.J. and Camacho, D. (2016), "Social big data: recent achievements and new challenges", *Information Fusion*, Vol. 28, pp. 45-59.

Blazquez, D. and Domenech, J. (2018), "Big data sources and methods for social and economic analyses", *Technological Forecasting and Social Change*, Vol. 130, pp. 99-113.

Buhalis, D. and Sinarta, Y. (2019), "Real-time co-creation and nowness service: lessons from tourism and hospitality", *Journal of Travel & Tourism Marketing*, Vol. 36 No. 5, pp. 563-582.

Buhalis, D., Harwood, T., Bogicevic, V., Viglia, G., Beldona, S. and Hofacker, C. (2019), "Technological disruptions in services: lessons from tourism and hospitality", *Journal of Service Management*, available at: https://dx.doi.org/10.1108/JOSM-12-2018-0398

Cox, M. and Ellsworth, D. (1997), "Managing big data for scientific visualization", *ACM Siggraph, MRJ/ NASA Ames Research Center*, Vol. 5, pp. 1-17.

Creswell, J.W. (2014), *Research Design: Qualitative, Quantitative and Mixed Method Approaches*, 4th Ed., Sage Publications, Thousand Oaks, CA.

Davenport, T.H. (2017), "How analytics have changed in the last 10 years", *Harvard Business Review*, June 2017.

Deng, N. and Li, X. (Robert). (2018), "Feeling a destination through the 'right' photos: a machine learning model for DMOs' photo selection", *Tourism Management*, Vol. 65, pp. 267-278.

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.

Fuchs, M., Höpken, W. and Lexhagen, M. (2016), *Big Data and Business Intelligence in the Travel and Tourism Industry*, Mid Sweden University, Östersund.

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.

Hilbert, M. and López, P. (2011), "The world's technological capacity to store, communicate, and compute information", *Science*, Vol. 332 No. 6025, pp. 60-65.

IBM (2016), "Big data and analytics", available at: www-01.ibm.com/software/data/bigdata/what-is-big-data.html (accessed 27 November 2016).

Liu, Y., Tseng, F.-M. and Tseng, Y.-H. (2018), "Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model", *Technological Forecasting and Social Change*, Vol. 130, pp. 123-134.

Mariani, M.M. and Borghi, M. (2018), "Effects of the Booking.com rating system: bringing hotel class into the picture", *Tourism Management*, Vol. 66, pp. 47-52.

Mariani, M.M. and Visani, F. (2019), "Embedding eWOM into efficiency DEA modelling: an application to the hospitality sector", *International Journal of Hospitality Management*, Vol. 80, pp. 1-12.

Mariani, M.M., Borghi, M. and Gretzel, U. (2019a), "Online reviews: differences by submission device", *Tourism Management*, Vol. 70, pp. 295-298.

Mariani, M.M., Borghi, M. and Kazakov, S. (2019b), "The role of language in the online evaluation of hospitality service encounters: an empirical study", *International Journal of Hospitality Management*, Vol. 78, pp. 50-58.

Mariani, M.M., Di Fatta, G. and Di Felice, M. (2019c), "Understanding customer satisfaction with services by leveraging big data: the role of services attributes and consumers' cultural background", *IEEE Access*, Vol. 7 No. 8580523, pp. 8195-8208.

Mariani, M.M., Mura, M. and Di Felice, M. (2018b), "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.

Mariani, M.M., Baggio, R., Fuchs, M. and Höpken, W. (2018a), "Business intelligence and big data in hospitality and tourism: a systematic literature review", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 12, pp. 3514-3554.

Mariani, M.M., Baggio, R., Buhalis, D. and Longhi, C. (Eds) (2013), "Introduction", *Tourism Management, Marketing, and Development: Volume I: The Importance of Networks and ICTs*, Palgrave, New York, pp. 1-11, doi: 10.1057/9781137354358_1.

Nambisan, S. (2017), "Digital entrepreneurship: toward a digital technology perspective of entrepreneurship", *Entrepreneurship Theory and Practice*, Vol. 41 No. 6, pp. 1029-1055.

Pauleen, D.J. and Wang, W.Y.C. (2017), "Does big data mean big knowledge? KM perspectives on big data and analytics", *Journal of Knowledge Management*, Vol. 21 No. 1, pp. 1-6.

Pigni, F., Piccoli, G. and Watson, R. (2016), "Digital data streams: creating value from the real-time flow of big data", *California Management Review*, Vol. 58 No. 3, pp. 5-25.

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. and Harnisch, M. (2015), *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*, Boston Consulting Group, Vol. 9.

Stylos, N. and Zwiegelaar, J. (2019), "Big data as a game changer: how does it shape business intelligence within a tourism and hospitality industry context?", in Sigala, M., Rahimi, R. and Thelwall, M. (Eds), *Big Data and Innovation in Tourism, Travel, and Hospitality*, Springer, pp. 163-181.

Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S. and Rojo-Álvarez, J.L. (2018), "Using big data from customer relationship management information systems to determine the client profile in the hotel sector", *Tourism Management*, Vol. 68, pp. 187-197.

The Economist (2017), "The world's most valuable resource is no longer oil, but data", available at: www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (accessed 5 June 2019).

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017), "A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism", *Tourism Management*, Vol. 58, pp. 51-65.

Zhao, Y., Xu, X. and Wang, M. (2019), "Predicting overall customer satisfaction: big data evidence from hotel online textual reviews", *International Journal of Hospitality Management*, Vol. 76, pp. 111-121.

## About the author

Marcello Mariani is Full Professor of Entrepreneurship and Management at the Henley Business School, University of Reading (United Kingdom) and member of the Henley Center for Entrepreneurship, the Academy of Management and the European Institute for Advanced Studies in Management. After earning a PhD in Business Administration at the University of Bologna (Italy), and a Post Doc at the University of Technology Sydney (Australia), he has been an Academic Faculty Member of the University of Bologna and acted as Visiting Professor at the Stern School of Business, New York University (USA) and a number of other universities. His current research interests include big data analytics, eWOM, digital business models, AI, automation and coopetition strategies in services industries. Marcello Mariani can be contacted at: m.mariani@henley.ac.uk

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com