

# Big Data, Business Intelligence and Tourism: a brief analysis of the literature

Rodolfo Baggio  
Master in Economics and Tourism  
Bocconi University, Milan, Italy  
rodolfo.baggio@unibocconi.it

paper presented at:

IFITTalk@Östersund: Big Data & Business Intelligence in the Travel & Tourism Domain.  
ETOUR, Mid-Sweden University, Östersund (SE), 11-12 April 2016

## 1 Problem Definition

Big Data is, today, a very popular buzzword (a Google search provides more than 55 million items containing the expression). The term identifies, as known, the massive volume of both structured and unstructured data apparently available on the Web and difficult to process using traditional software techniques or by using traditional statistical methods. It is a rapidly emerging field of inquiry, often hailed as a crucial factor for increasing economic prosperity and understanding or resolving societal problems (Mayer-Schönberger & Cukier, 2013).

Big Data (BD) are considered by many an incredible opportunity for its supposed capacity to provide answers to practically any question that could be asked about people's behaviours, views and feelings. As a matter of fact, it is rather surprising to see that a phenomenon once considered causing puzzlement and confusion, the so called *information overload*, once changed name into *Big Data* is now believed to be a kind of silver bullet, able to provide a wealth of valuable and unquestionable insights into many aspects of the modern life of individuals, organisations and markets (Mayer-Schönberger & Cukier, 2013; McAfee et al., 2012).

Many of these claims, though, look more than reasonable and, actually, the capability to examine complex phenomena by combining so widely available sources of information can be a remarkable advantage for those who can fully exploit them (Bedeley & Nemati, 2014).

On the other hand, BD present a good number of challenges and risks, well discussed in a number of works (Boyd & Crawford, 2012; Fan et al., 2014; McFarland & McFarland, 2015). They mainly refer to the technical and methodological difficulties in treating so large volumes of rapidly changing sets of data. Besides that, there is a need for a good set of specialised skills and resources, and a different approach is deemed necessary with respect to the one that for the last centuries has characterised the collection and the analysis of data (Chen et al., 2014).

Nonetheless, scholars and practitioners overall agree that there are remarkable benefits in having access to a vast amount of data that cover practically any aspect of human life, mainly because they are "spontaneously" generated so do not suffer from selection biases that can be present in traditional investigation methods. In any case, even taking into account the methodological issues, BD can be a useful and important complement to more stable, rigorous research methods (Kitchin & Lauriault, 2015).

Big Data have also started to be a source for Business Intelligence (BI) activities. The tradition of BI analytics is longer, but the field is very sensitive to all data and information sources that can provide a better return on the investment. Therefore both subjects are highly complementary. Advanced analytics and better and richer sources can provide a deeper perspective on the data, that can benefit from more structured and rigorous experience. The interpretation layer provided by business intelligence can thus be crucial to making advanced BD analytics actionable (Liebowitz, 2013).

In the last years, tourism has widely recognised the need for a more *customer-focused* approach, that primarily values tourists' needs, preferences and requirements in order to increase the goodness of their experience and achieve a better satisfaction, that turns out to be an important determinant in all travel choices and decisions (Correia et al., 2013; Prayag et al., 2013).

Given these premises, the question is: to what extent is tourism academia aware of, and is working on these subjects?

To answer, at least partially, this question this short note presents the results of an analysis of the recent literature on Big Data and business intelligence and the application of the related techniques to the field of travel, tourism, hospitality and leisure.

## 2 Materials and Methods

The source for this work is the Scopus database. Although, obviously, not a complete source of academic works, with its more than 20,000 titles from about 5,000 international publishers, it can be well considered one of the most comprehensive repository of the world's research output across a wide range of fields. To this we add the IFITT digital library (<http://www.ifitt.org/resources/digital-library/>) which indexes 936 works published in the Journal of Information Technology and Tourism and in the proceedings of the Enter conferences.

A search with "Big Data" in the titles, abstracts and keywords returns 14,051 works. The time distribution is largely uneven and testifies the very recent growth of interest in the subject. A look at Fig. 1 (works published in the last 15 years) shows an almost exponential growth, with an acceleration in the last five years.

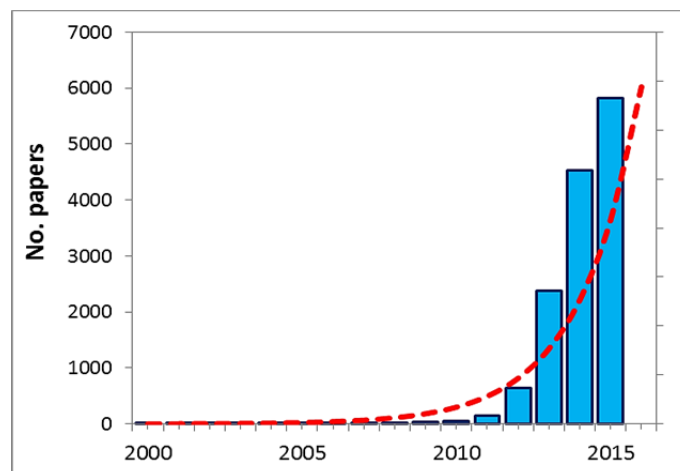


Fig. 1 Time distribution of Big Data works published in the last 15 years

The situation with BI is different. Here we find 16,496 works (with "Business Intelligence" in the titles, abstracts and keywords) with a much wider distribution over time and a moderate growth in the last 15 years (Fig. 2).

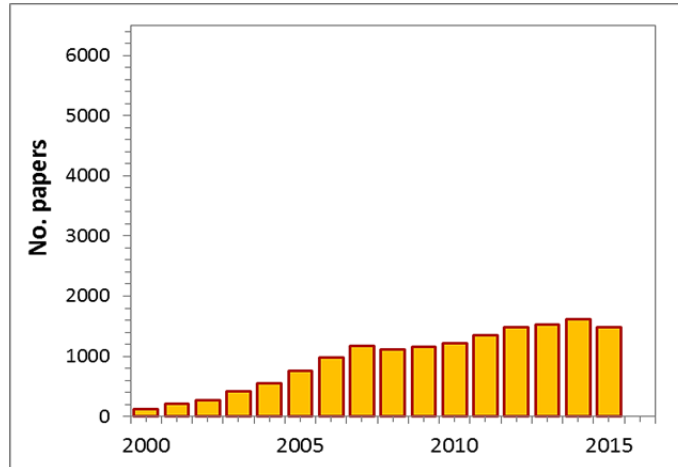


Fig. 1 Time distribution of Business Intelligence works published in the last 15 years

The works related to the wide area of tourism have been identified by restricting the search to these works having “*travel, tourism, tourist, hospitality or leisure*” in the titles, abstracts and keywords. Of these, then, the papers published in tourism and hospitality journals have been selected (Scopus indexes about 80 journals in the field).

Finally, titles and abstracts have been manually inspected to further pick out the works actually dealing with BD and BI.

### 3 Results and Discussion

The total number of BD works in the area of tourism selected as described above is 127. the BI works are 529. Their time distribution is given in Fig. 3 (for the last five years).

Incredibly, it seems that besides the much hype about the BD issue, not many tourism researchers have decided to pay some effort in studying these topics, and only a handful of them have invested time and resources in considering the possibilities of an application of Big Data to the tourism and hospitality field. The BI field, instead, can count on a relatively higher throughput, even if still of a very limited size (at least with respect to the total production).

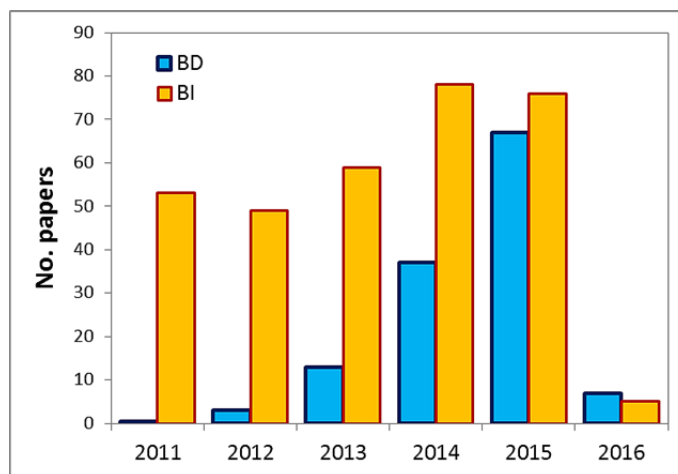


Fig. 3 Time distribution of BD and BI tourism related works for the last five years

What is more *interesting* is the fact that only 20 of these BD works appear in tourism or hospitality journals, and therefore more accessible to the tourism academic community. Same situation is found for the BI papers: only 18 are available in tourism publications. All the others come from Computer Science (mainly), Transportation, Marketing Management or Geography publications.

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

The first thing to notice, for what concerns Big Data, is that the abstracts and titles contain rather *generic* terms, with no or little reference to the specific terminology often used in works about Big Data (see Fig. 3).



Fig. 3 A word cloud with the most used terms in the BD papers selected

Some papers present a roughly general discussion about Big Data or about the importance of using Big Data for improving and extending present research activities (Buhalis & Foerste, 2015; Dolnicar & Ring, 2014; Wang et al., 2015).

Despite the call for a better integration between official statistics and Big Data (see e.g. Heerschap et al., 2014; Lam & McKercher, 2013), not many attempt to find a solution. Yang et al. (2014) use web traffic volume data of a destination marketing organisation to predict hotel demand, showing an improvement in the error reduction more traditional forecasting models, and 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. Fuchs et al. (2014) show how BD analytics can be beneficial for BI practices in a tourism destination and propose an architectural solution that combines the different sources of data.

Advanced approaches such as machine learning techniques, artificial intelligence 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. 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 employing is the paper by employing these advanced techniques (support vector machines, naïve Bayes classifiers, latent semantic indexing etc.)

Not many other papers actually use online sources. This is the case of Xiang et al. (2015) that analyse a large corpus of tourists' reviews and derive a number of interesting considerations about hotel guest experience and its association with satisfaction ratings, or Marine-Roig et al.



One interesting factor emerge: among the most recent tourism BI works, four (actually 22%) are also catalogued in the BD listings (Fuchs et al., 2014; Lam, et al., 2013; Marine-Roig et al., 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, or should be used, for enriching the business intelligence practices of destinations and operators.

#### 4 Concluding Remarks

Although quite popular and strongly pushed by many, the idea that Big Data can be a very useful source of information for the tourism sector seems to be still a bit ignored by the researchers in the field, at least when “real” work is concerned. Same situation seems to happen for what concerns business intelligence studies. It is difficult to understand the reasons for this situation.

A hint may be that the resources (hardware and software) needed to actually treat huge quantities of data are not normal available to tourism researchers, but rather sit in computer science departments, and 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. For BI, the same could be said, as good practices call for well and rationally designed, organised and managed information systems.

These issues could be solved by establishing good interdisciplinary collaborations. Moreover, for the future generations, it is important to start introducing well-tailored educational programs in the tourism studies curricula.

#### References

- Barbieri, C., & Sotomavor, S. (2013). Surf travel behavior and destination preferences: An application of the Serious Leisure Inventory and Measure. *Tourism Management*, 35, 111-121.
- Bedeley, R., & Nemati, H. (2014). *Big Data Analytics: A Key Capability for Competitive Advantage*. Paper presented at the 20th Americas Conference on Information Systems (AMCIS), Savannah, GA, 7-9 August.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Buhalis, D., & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. *Journal of Destination Marketing & Management*, 4(3), 151-161.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Correia, A., Kozak, M., & Ferradeira, J. (2013). From tourist motivations to tourist satisfaction. International. *Journal of Culture, Tourism and Hospitality Research*, 7(4), 411-424.
- d'Amore, M., Baggio, R., & Valdani, E. (2015). A practical approach to big data in tourism: a low cost Raspberry Pi cluster. In I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015 (Proceedings of the International Conference in Lugano, Switzerland, February 3-6)* (pp. 169-181). Berlin - Heidelberg: Springer.
- Dolnicar, S., & Ring, A. (2014). Tourism marketing research: Past, present and future. *Annals of Tourism Research*, 47, 31-47.
- Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, 1(2), 293-314.
- Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework for tourism sustainability: A business intelligence application from Sweden. *Turizam: znanstveno-stručni časopis*, 61(2), 121-148.
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management*, 3(4), 198-209.
- Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103-114.

- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3), 260-271.
- Heerschap, N., Ortega, S., Priem, A., & 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, Prague, CZ, 15-16 May. Retrieved July 2014 from [http://www.tsf2014prague.cz/assets/downloads/Paper%201.2 Nicolaes%20Heerschap NL.pdf](http://www.tsf2014prague.cz/assets/downloads/Paper%201.2%20Nicolaes%20Heerschap%20NL.pdf).
- Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. *Information Technology & Tourism*, 15(2), 101-130.
- Kasahara, H., Mori, M., Mukunoki, M., & Minoh, M. (2015). Transportation Mode Annotation of Tourist GPS Trajectories Under Environmental Constraints. In A. Inversini & R. Schegg (Eds.), *Information and Communication Technologies in Tourism 2015* (pp. 523-535). Heidelberg: Springer.
- Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, 80(4), 463-475.
- Köseoglu, M. A., Ross, G., & Okumus, F. (2016). Competitive intelligence practices in hotels. *International Journal of Hospitality Management*, 53, 161-172.
- Lam, C., & McKercher, B. (2013). The tourism data gap: The utility of official tourism information for the hospitality and tourism industry. *Tourism Management Perspectives*, 6, 82-94.
- Liebowitz, J. (Ed.). (2013). *Big data and business analytics*. Boca Raton, FL: CRC Press.
- Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, 54, 321-343.
- Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162-172.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big Data. The management revolution. *Harvard Business Review*, 90(10), 61-67.
- McFarland, D. A., & McFarland, H. R. (2015). Big Data and the danger of being precisely inaccurate. *Big Data & Society*, 2(2), 1-4.
- Menner, T., Höpken, W., Fuchs, M., & Lexhagen, M. (2016). Topic detection – Identifying relevant topics, within touristic UGC. In A. Inversini & R. Schegg (Eds.), *Information and Communication Technologies in Tourism 2015* (pp. 411-423). Heidelberg: Springer.
- Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2014). Tracing Tourists by Their Digital Footprints The Case of Austria. *Journal of Travel Research*, doi: 10.1177/0047287514563985.
- Park, S. B., Ok, C. M., & Chae, B. K. (2015). Using Twitter Data for Cruise Tourism Marketing and Research. *Journal of Travel & Tourism Marketing*, doi: 10.1080/10548408.10542015.11071688.
- Prayag, G., Hosany, S., & Odeh, K. (2013). The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions. *Journal of Destination Marketing & Management*, 2(2), 118-127.
- Qiao, X., Zhang, L., Li, N., & Zhu, W. (2014). Constructing a Data Warehouse Based Decision Support Platform for China Tourism Industry. In Z. Xiang & I. Tussyadiah (Eds.), *Information and Communication Technologies in Tourism 2014* (pp. 883-893). Heidelberg: Springer.
- Ritchie, R. J., & Ritchie, J. B. (2002). A framework for an industry supported destination marketing information system. *Tourism Management*, 23(5), 439-454.
- RocaSalvatella. (2014). *Big Data and Tourism: New Indicators for Tourism Management*. Barcelona: Telefónica I+D and RocaSalvatella. Retrieved September, 2015, from [http://www.rocasalvatella.com/sites/default/files/big\\_data\\_y\\_turismo-eng-interactivo.pdf](http://www.rocasalvatella.com/sites/default/files/big_data_y_turismo-eng-interactivo.pdf).
- Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2014). Sentiment Analysis – Implementation and Evaluation of Methods for Sentiment Analysis with Rapid-Miner®. In Z. Xiang & I. Tussyadiah (Eds.), *Information and Communication Technologies in Tourism 2014* (pp. 253-265). Heidelberg: Springer.
- Wang, X. L., Yoonjung Heo, C., Schwartz, Z., Legohérel, P., & Specklin, F. (2015). Revenue management: progress, challenges, and research prospects. *Journal of Travel & Tourism Marketing*, 32(7), 797-811.
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.
- Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433-447.