

BIG DATA PARA A GERAÇÃO DE INDICADORES DE TURISMO: FONTES DE DADOS E APLICAÇÕES

Alexandre Augusto Biz¹
João Artur Souza²
Rafael Bassegio Caumo³

RESUMO

O artigo tem por objetivo sintetizar a produção científica sobre o uso de dados do tipo Big Data na geração de indicadores de turismo no que concerne a possíveis fontes de dados, aplicações práticas e principais agentes envolvidos com a pesquisa nesta área. Para tanto, foi conduzida uma pesquisa bibliográfica sob a forma de revisão sistemática de literatura em 175 artigos coletados no dia 23 de abril de 2019 nas bases Scopus e Web of Science. Durante a análise, percebeu-se predominância de produção científica europeia, asiática e norte-americana que se utiliza de fontes de dados presentes na World Wide Web para estimação do contingente e sentimento dos turistas. Espera-se que os resultados possam auxiliar produtores e usuários de conhecimento em turismo a vislumbrarem métodos alternativos inovadores, mais econômicos, de geração de indicadores sobre o setor.

PALAVRAS-CHAVE: Big Data. Turismo. Indicadores. Conhecimento. Inovação.

1. Doutor, e-mail: alexandre.biz@ufsc.br
2. Doutor, e-mail: jartur@gmail.com
3. Doutorando, e-mail: rbcaumo@gmail.com

1 INTRODUÇÃO

Indicadores são instrumentos essenciais para os processos de planejamento, gestão e monitoramento do turismo, fornecendo informação acurada para momentos de tomada de decisão (UNWTO, 2007; UN, 2010). Desta forma, ao subsidiarem ações em busca de resultados concretos, podem ser considerados representações explícitas de conhecimento (SCHREIBER et al., 2000).

A produção de indicadores de turismo, atrelada predominantemente às organizações produtoras de estatísticas públicas, se dá tradicionalmente por meio de censos e de pesquisas amostrais e pela utilização de registros administrativos. Os dois primeiros, ainda dominantes, possuem reconhecido rigor científico (IBGE, 2013), mas, pelo fato de geralmente considerarem levantamentos em campo intermediados por agentes humanos, dependem de processos muitas vezes caros, demorados e de operacionalização complexa (CITRO, 2014; HACKL, 2016).

Entretanto, a revolução dos dados (KITCHIN, 2014) e o Big Data (UNECE, 2013) passaram a oferecer possibilidades com características de inovação de ruptura para o contorno a uma série de problemáticas associadas ao modelo tradicional de produção de estatísticas (CITRO, 2014). Organizações produtoras que não atentarem para tais possibilidades de inovação deverão perder espaço no mercado das estatísticas públicas – até então monopolizado – uma vez que novos players estarão em condições de gerar

indicadores socioeconômicos com baixo custo, em alto nível de detalhamento e oportunismo temporal (DEMUNTER, 2017).

Nesse contexto, a presente pesquisa se propõe a sintetizar a produção científica sobre o uso de dados do tipo Big Data na geração de indicadores de turismo no que concerne a possíveis fontes, aplicações práticas e principais agentes envolvidos com a pesquisa nesta área. Para tanto, foi conduzida uma pesquisa bibliográfica sob a forma de revisão sistemática em coleta realizada no dia 23 de abril de 2019 nas bases Scopus e Web of Science. Espera-se que o conteúdo aqui explorado permita que produtores e usuários de conhecimento em turismo vislumbrem métodos alternativos inovadores, mais econômicos, de geração de indicadores sobre o setor.

O artigo está organizado em cinco seções, incluindo-se esta primeira, introdutória ao tema e ao propósito da pesquisa. A segunda traz um referencial teórico, com contextualização da problemática e de importantes conceitos que necessitam ser demarcados. A terceira seção apresenta a metodologia aqui empregada. A quarta apresenta e analisa os resultados no que concerne aos principais envolvidos com o tema, fontes de Big Data em utilização e aplicações práticas que vêm sendo realizadas. Por fim, a última seção tece considerações finais.

2 REFERENCIAL TEÓRICO

2.1 Conhecimento e inovação

Conhecimento e inovação são conceitos transversais que perpassam a presente pesquisa e por isso merecem definição preliminar.

Por conhecimento, entende-se aqui conteúdo ou processo efetivado por agentes humanos ou artificiais em atividades de geração de valor científico, tecnológico, econômico, social ou cultural (PACHECO, 2014). Enquanto insumo para a cadeia de geração de valor, já foi amplamente percebido como um importante fator de produção (DRUCKER, 1993). Um ativo organizacional que, na visão epistemológica dos cognitivistas, não existe somente na forma tácita, mas pode também ser explicitado, modelado e inserido em sistemas computacionais (VENZIN et al., 1998; SCHREIBER et al., 2000; PACHECO, 2016).

Inovação, por sua vez, de acordo com o *Manual de Oslo* (OECD, 2005), pode ser vista como a implementação de um produto (bem ou serviço) novo ou significativamente melhorado, ou um processo, ou um novo método de marketing, ou um novo método organizacional nas práticas de negócios, na organização do local de trabalho ou nas relações externas. Nesta visão, o produto ou processo não precisa necessariamente ser inédito no mundo para ser considerado inovador, porém deve o ser no mercado ou na empresa onde está sendo aplicado. No Brasil, em âmbito oficial, de acordo com a Lei 10.973 de 2 de dezembro de 2004, uma definição genérica estendida do *Manual de Oslo* é adotada e diz que inovação é a introdução de novidade ou aperfeiçoamento no ambiente produtivo ou social que resulte em novos produtos, processos ou serviços.

2.2 Indicadores de turismo

Um indicador é, geralmente, uma medida quantitativa utilizada para ilustrar e comunicar um conjunto de fenômenos complexos de uma forma simples (OCDE, 2002). Quando tratam de aspectos econômicos, demográficos, sociais e ambientais de uma nação, são ditos indicadores socioeconômicos, insumos fundamentais para o planejamento e a formulação de políticas e estratégias no mundo contemporâneo (JANUZZI; GRACIOSO, 2002; SCHNORR-BACKER, 2016).

Sob a ótica da engenharia do conhecimento dos cognitivistas (VENZIN et al., 1998; Pacheco, 2016), dados e informações colocados em prática na realização de tarefas também são considerados conhecimento (SCHREIBER et al., 2000). Dessa forma, indicadores socioeconômicos podem ser considerados representações explícitas de conhecimento quando apoiarem ações em busca por resultados concretos.

No âmbito do turismo, a visão sobre a utilidade de indicadores é a mesma. Segundo UNWTO (2007), trata-se de instrumentos essenciais para os processos de planejamento, gestão e monitoramento do turismo, fornecendo informação acurada para os processos de tomada de decisão. Mais especificamente, são necessários para a formulação de estratégias de marketing, o fortalecimento das relações entre instituições, a avaliação da eficiência e da efetividade das decisões gerenciais e a mensuração do turismo no âmbito da economia nacional (UN, 2010).

Por vezes, conforme faz Unwto (2017), indicadores de turismo são classificados em referentes ao: (i) turismo de entrada; (ii) turismo doméstico; (iii)

turismo de saída; (iv) indústria do turismo; (v) empregos; (vi) complementares (macroeconômicos). Outra tipologia bastante comum trabalha sob uma ótica ligeiramente diferente, organizando tais categorias de modo que os indicadores sejam agrupados em referentes às perspectivas da demanda e da oferta turística (EUROSTAT, 2014; UN, 2010; BAGGIO, 2019).

A produção destes indicadores, por sua vez, tem sido tradicionalmente conduzida por organizações produtoras de estatísticas públicas utilizando-se principalmente: de censos ou pesquisas amostrais conduzidas por intermediários humanos por meio da aplicação de questionários; e de registros administrativos presentes em sistemas oficiais ou de empresas privadas prestadoras de serviços turísticos (UNWTO, 2017; EUROSTAT, 2014; UN, 2010; BAGGIO, 2019; OLIVEIRA; BARACHO, 2018). O primeiro grupo, ainda predominante, reconhecidamente rigoroso em termos científicos (IBGE, 2013), possui processos em geral caros, demorados e de operacionalização complexa (CITRO, 2014; HACKL, 2016).

2.3 Big Data e a revolução dos dados

Há algum tempo, se percebe uma avalanche de dados sendo criados diariamente (MILLER, 2010; HELBING et al., 2016), impulsionados pela popularização da internet, pelo avanço das tecnologias da informação e pela consolidação da era digital (SCHMIDT; COHEN, 2013). Tais dados, ditos dados digitais, derivam dos rastros – tanto inconscientes quanto deliberados – da atividade humana e não humana capturada por dispositivos inseridos no universo dos sensores, scanners e da internet das coisas, dos smartphones e aplicativos, das páginas da internet, dos mecanismos de buscas, das redes

sociais virtuais, dos blogs e web fóruns, entre outros. Por vezes, o termo Big Data surge como sinônimo desses dados digitais, como faz Unece (2013) ao apresentar e classificar as possíveis fontes desse tipo de dado.

Nesse contexto, se no passado a predominância por dados no mundo estava vinculada a dados ditos *small data*, com características de analógicos, escassos e de acesso limitado, hoje a quase totalidade dos dados produzidos é digital, gerados em vasta escala, com velocidade e em variedade de domínios e formatos (KITCHIN, 2015). Esta transformação compõe o processo do que é por Kitchin (2014) chamado de revolução dos dados.

Antes mesmo de ter seu conceito formalizado, este processo já era percebido como tendo características de inovação de ruptura ao desafiar o *statu quo* de como os dados são produzidos, geridos, analisados, armazenados e utilizados, modificando a maneira como o conhecimento é produzido, os negócios são conduzidos e a governança é promulgada. Há alguns anos, Mayer-Schönberger e Cukier (2013) já diziam que o Big Data se tratava de um fenômeno que estaria pronto para “chacoalhar” tudo, dos negócios e das ciências até os sistemas de saúde, os governos, a educação, a economia, as humanidades e todos os outros aspectos da sociedade.

Os produtores de estatísticas públicas e/ou oficiais – disponibilizadas na maioria das vezes sob a forma de indicadores socioeconômicos – também já atentaram para as oportunidades trazidas pelo Big Data e a revolução dos dados (UNSC, 2014; EUROSTAT, 2013). Vislumbram possibilidades de contorno de uma série de problemáticas associadas ao modelo tradicional de produção de estatísticas públicas (CITRO, 2014; HACKL, 2016).

Dentre os possíveis domínios das estatísticas públicas de cunho socioeconômico, o turismo está na linha de frente das inovações associadas ao Big Data no que concerne a novas fontes e métodos (DEMUNTER, 2017). Organizações produtoras que não atentarem para tais possibilidades de inovação deverão perder espaço no mercado das estatísticas públicas – até então monopolizado – uma vez que novos players

estarão em condições de gerar indicadores socioeconômicos com baixo custo, em alto nível de detalhamento e oportunismo temporal (DEMUNTER, 2017). Assim, a prática de inovar se fará imprescindível, pois poderá representar fator crucial de sucesso e sobrevivência em meio a um novo cenário globalizado e competitivo (PLENTZ et al., 2015; TIDD et al., 2008; BROWN; EISENHARDT, 1995).

3 METODOLOGIA

Este estudo se propõe a ser uma pesquisa científica no âmbito das ciências empíricas sociais, conforme definições trazidas por Gil (2008). Em relação ao alcance do objetivo de sintetizar a produção científica sobre o uso de dados do tipo Big Data na geração de indicadores de turismo, o presente estudo almeja alcançar o nível exploratório – objetivando proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito –, com delineamento do tipo bibliográfico – elaborado a partir de materiais já publicados (GIL, 2008).

O levantamento bibliográfico é realizado de forma sistemática (FORBES, 1998), partindo de uma pergunta claramente formulada e utilizando métodos sistemáticos e explícitos para: identificar, selecionar e avaliar criticamente

pesquisas relevantes; e coletar e analisar dados dos estudos (GREEN; HIGGINS, 2005), com critérios para a busca que seguem o método PRISMA (MOHER et al., 2009). Ao final, os documentos coletados foram analisados em busca de conteúdos que permitam responder ao objetivo central do artigo – com identificação: dos principais agentes envolvidos com a pesquisa sobre o tema; de possíveis fontes de dados a serem utilizadas; e de aplicações práticas que já vêm sendo propostas e/ou trabalhadas.

Os documentos foram selecionados a partir de busca que percorreu as bases apresentadas no Quadro 1 – realizada em 23 de abril de 2019. As bases Scopus e Web of Science foram escolhidas por sua reconhecida cobertura de literatura científica revisada (*peer-reviewed*).

Quadro 1: Informações sobre a origem dos documentos selecionados

Base	Tipo de busca	Retornos
Scopus (Elsevier)	Título, resumo e palavras-chave	157 documentos
Web of Science (Clarivate Analytics)	Tópico	99 documentos

Fonte: Dos autores (2019)

O critério de busca selecionou documentos que contivessem pelo menos uma das expressões de cada um dos dois blocos apresentados a seguir. Dessa forma, a busca percorreu $6 \times 15 = 90$ possíveis combinações das expressões contempladas pelos dois blocos.

- **Bloco 1:** *Tourism statistics; tourism data; tourism indicators; tourism information; touristic information; tourism research.*

- **Bloco 2:** *Big data; multisource; data science; digital data; social media; facebook; Instagram; social data; personal data; crowdsourcing; crowdsensing; social network; twitter; smart destination; smart tourism destination.*

Dos 256 documentos que apresentaram as buscas nas diferentes bases, 75 eram repetidos e 6 correspondiam ao prefácio de anais de eventos. Esses 81 foram excluídos, fazendo com que a quantidade final de documentos exclusivos considerados neste estudo fosse de 175.

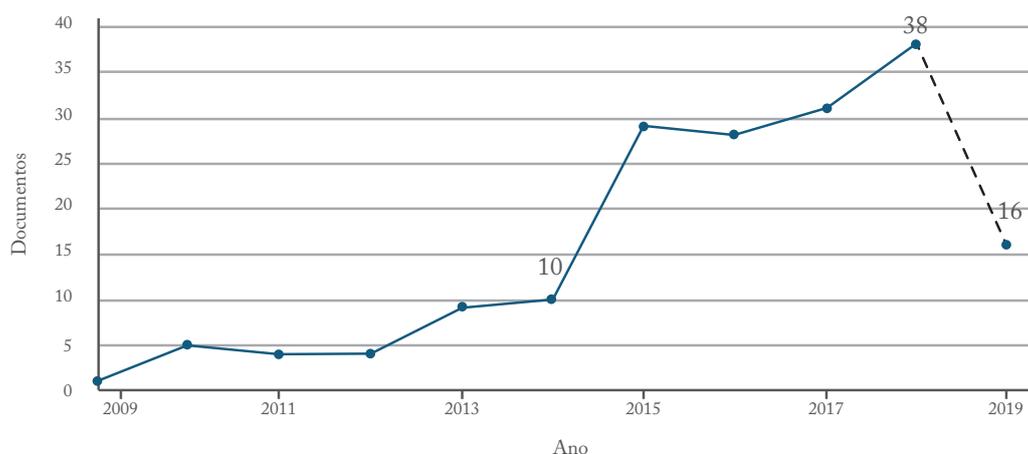
4 RESULTADOS

4.1 Resumo bibliométrico

A evolução temporal da produção absoluta dos materiais que compõem a base técnico-científica aqui analisada é apresentada no Gráfico 1. Destaca-se a atualidade e o crescimento do tema, baseando-se no fato de que a publicação mais antiga capturada é de 2009 e que a quantidade de produções segue uma tendência

de crescimento desde então. Percebe-se, por exemplo, uma quantia de publicações em 2018 (38) que é aproximadamente quatro vezes a quantia de publicações de 2014 (10), quatro anos antes. A quantidade encontrada em 2019 está incompleta uma vez que a busca foi realizada em 23 de abril de 2019.

Gráfico 1: Evolução temporal absoluta dos documentos analisados, por ano



Fonte: Dos autores (2019) a partir do levantamento bibliográfico

Nota: O quantitativo de 2019 corresponde ao encontrado na data da busca (23/04/2019).

Como forma de análise dos pesquisadores e instituições mais engajados com o assunto, foram filtrados todos os autores que apareceram em mais de dois documentos analisados. Os resultados, com nome do autor, quantidade de publicações em que aparece como autor e país e nome da instituição de vínculo especificada nos documentos, estão apresentados no Quadro 2.

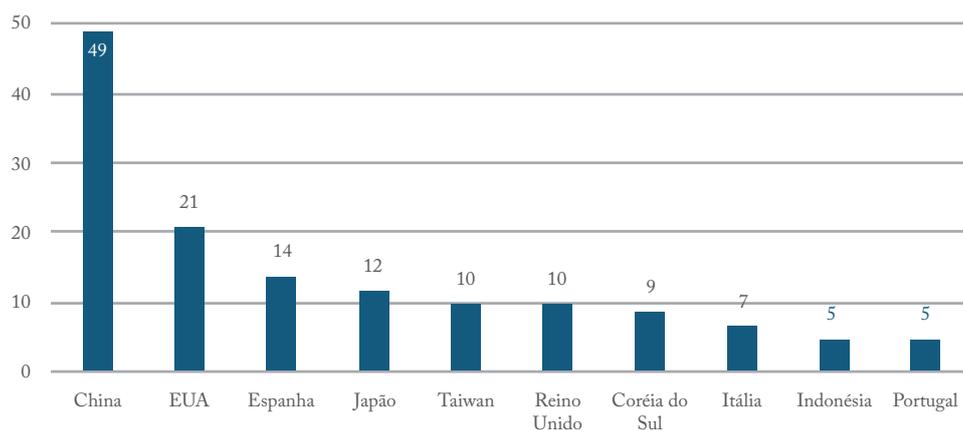
Quadro 2: Informações sobre a origem dos documentos selecionados

Autor	Qtd.	País	Instituição
Rob Law	4	Hong Kong	School of Hotel & Tourism Management (The Hong Kong Polytechnic University)
Rodolfo Baggio	3	Itália; Rússia	Master in Economics and Tourism (Bocconi University); National Research Tomsk Polytechnic University
Tsutomu Endo	3	Japão	Kyushu Institute of Technology
Bin Guo	3	China	School of Computer Science (Northwestern Polytechnical University)
Tong Guo	3	China	School of Computer Science (Northwestern Polytechnical University)
Chaang-luan Ho	3	Taiwan	Department of Leisure Services Management (Chaoyang University of Technology)
Shunsuke Inoue	3	Japão	Kyushu Institute of Technology
Kazutaka Shimada	3	Japão	Kyushu Institute of Technology

Fonte: Dos autores (2019) a partir do levantamento bibliográfico

As informações trazidas pelo Quadro 2 sugerem uma produção predominantemente asiática e europeia. O Gráfico 2 – que apresenta a distribuição da produção por países vinculados a pelo menos 5 dos 175 documentos – acrescenta os Estados Unidos na relação dos principais países produtores de conhecimento sobre o tema. China (1º) e Espanha (3º) completam a lista dos três principais. Destaca-se que nenhum país sul ou central-americano apareceu como associado a mais de uma publicação, sugerindo uma baixa produção científica em tais localidades.

Gráfico 2: Distribuição absoluta dos documentos analisados, por principais países produtores



Fonte: Dos autores (2019) a partir do levantamento bibliográfico

No que concerne às fontes, os periódicos que mais contemplam documentos selecionados pela busca sistemática aqui realizada estão apresentados no Quadro 3.

Quadro 3: Distribuição absoluta dos documentos analisados, por principais periódicos de origem

Periódico	Quantidade
<i>Current Issues in Tourism</i>	9
<i>Asia Pacific Journal of Tourism Research</i>	5
<i>ACM International Conference Proceeding Series</i>	4
<i>Lecture Notes in Computer Science</i>	4
<i>Tourism Review</i>	4
<i>Boletín de Estadística e Investigación Operativa</i>	3
<i>International Journal of Tourism Research</i>	3
<i>Journal of China Tourism Research</i>	3
<i>Journal of Sustainable Tourism</i>	3
<i>Journal of Travel and Tourism Marketing</i>	3
<i>Tourism Management</i>	3

Fonte: Dos autores (2019) a partir do levantamento bibliográfico

Por fim, a Figura 1 apresenta, de maneira visual, as palavras-chave listadas nas publicações selecionadas, dimensionadas proporcionalmente à quantidade de aparições.

Fonte de dados	Aplicação	Publicação
Beacons	Tendências e interesses dos turistas.	Guarda et al. (2018)
ChinaZ.com	Estrutura da rede das 120 principais páginas chinesas de informação turística da web.	Feng et al. (2015)
Dados de posicionamento de telefones móveis	Impacto econômico do turismo.	Berzina e Lauberte (2018)
	Quantidade de visitas ao país.	Pramana et al. (2017)
	Quantidade de turistas estrangeiros e suas movimentações.	Setiadi e Uluwiyah (2017)
Flickr	Avaliação, reconhecimento e classificação de pontos turísticos (foco na identificação de "Anabas").	Kitayama (2016)
	Quantidade e distribuição espacial de turistas (demanda).	Önder et al. (2014)
TripAdvisor	Satisfação, perfil dos turistas e períodos relevantes de visita de municípios e atrações em Minas Gerais.	Oliveira e Baracho (2018)
	Satisfação e opinião dos turistas.	Guo et al. (2016b)
Twitter	Motivações, percepções, preocupações e sentimentos daqueles que já realizaram e dos que consideram realizar turismo médico.	Claster et al. (2015)
	Padrões de movimentações (rotas entre atrações) dos turistas.	Hu et al. (2018)
	Sentimento/polaridade de manifestações de turistas.	Menchavez e Espinosa (2015); Shimada et al. (2011)
	Comportamento e interesses dos turistas.	Nozawa et al. (2017)
	Redes de fluxos turísticos entre países.	Provenzano et al. (2018)
SinaWeibo(China)	Quantidade de chegadas e origem dos turistas.	Liu e Shi (2019)
	Relação entre a popularidade de atrações turísticas e as multidões que de fato atraem.	Shi et al. (2017)
	Quantidade de turistas estrangeiros.	Yang et al. (2017)
Sistema proprietário de reservas on-line	Comportamento e atributos do turista que retorna ao hotel.	Park et al. (2018)

Fonte de dados	Aplicação	Publicação
Wi-fi logs	Movimentações dos turistas.	Ichifuji et al. (2016)
	Quantidade de turistas em tempo real.	Koide et al. (2016)
Wikipedia	Quantidade de turistas que saem do Reino Unido por mês.	Alis et al. (2015)
	Quantidade de turistas.	Donovan et al. (2017)

Fonte: Dos autores (2019) a partir do levantamento bibliográfico

Destaca-se que, para os fins desta investigação, a maioria das aplicações percebidas utilizam-se de fontes presentes na World Wide Web (páginas, redes sociais, blogs, mecanismos de busca, entre outros). Do total de aplicações, o percentual com fonte correspondente a sensores e internet das coisas (*beacons* e pontos de *wi-fi*), scanners (registros de pagamentos eletrônicos), smartphones (dados de posicionamento e registros de ligações) e outros dispositivos digitais não classificados como WWW é de aproximadamente 29%. Desta parcela, metade envolve a utilização de dados oriundos de serviços de telefonia.

Das fontes presentes na WWW, chama atenção a não percepção da utilização da rede social Instagram como fonte de dados. A hipótese aqui especulada é de que a política de termos e dados

fortemente restritiva do Instagram impeça uma série de coletas e acabe por afastar interessados. A rede Facebook, por sua vez, apareceu como fonte em apenas uma publicação, sugerindo que fenômeno semelhante possa estar no cerne da justificativa. Em sentido oposto, a rede Twitter foi a fonte presente na WWW com maior quantidade de aparições.

No que concerne às aplicações, as mais frequentes são aquelas que envolvem a estimação ou a projeção da quantidade ou demanda de turistas de entrada ou saída e aquelas que trabalham com satisfação, opinião, sentimentos e interesses manifestos. Ambas aparecem em aproximadamente 36% das publicações. Outras finalidades de aplicações encontradas estão apresentadas no Quadro 5.

Quadro 5: Percentual de publicações classificadas nos tipos de finalidade da aplicação

Finalidade	Publicações
Quantidade e/ou distribuição espacial dos turistas	35,7%
Satisfação, opiniões, sentimentos e interesses manifestos	35,7%
Fluxos turísticos, movimentações, atrações e locais visitados, tempo de estadia e gastos	21,4%
Perfil, atributos e origem dos turistas	10,7%
Impacto econômico	3,6%
Outros	10,7%

Fonte: Dos autores (2019) a partir do levantamento bibliográfico

Outras publicações não propuseram ou relataram experiências próprias de utilização de Big Data para a construção de indicadores e estatísticas de turismo, mas citaram finalidades e aplicações realizadas por terceiros. É o caso das revisões de literatura cujo objetivo está apresentado no Quadro 6.

Quadro 6: Revisões de literatura de temas correlatos ao proposto por este artigo

Publicação	Objetivo
Centobelli e Ndou (2019)	Análise de 109 artigos sobre Big Data no turismo, com identificação dos possíveis impactos gerados no setor, funcionalidades, tecnologias e metodologias.
Li et al. (2018)	Análise de 165 artigos para identificação de fontes e aplicações de Big Data no turismo, destacando tipos de dados e técnicas analíticas.
Liang et al. (2017)	Análise de 57 artigos com foco na percepção de tecnologias da informação e comunicação que vêm sendo adotadas no setor turístico Chinês.
Mariani et al. (2018)	Análise de 96 artigos sobre Big Data no turismo para identificação de aplicações práticas com fontes e tipos de dados e métodos de coleta, análise e visualização utilizados.
Shoval e Ahas (2016)	Análise de 45 artigos com foco na percepção de tecnologias utilizadas para rastreamento de turistas.
Thelwall (2019)	Revisão das principais abordagens de análise de sentimentos com foco em descrições práticas de como os métodos funcionam e podem ser aplicados no turismo.

Fonte: Dos autores (2019) a partir do levantamento bibliográfico

Além das aplicações e finalidades apresentadas nos Quadros 4 e 5 e nas revisões de literatura apresentadas no Quadro 6, cabe destacar que outras duas utilizações de Big Data para o setor turístico foram amplamente percebidas: as construções de sistemas customizados de informação e de recomendação (BIN et al., 2018; BOULAALAM et al., 2018; CHEN; CHEN, 2017; DEZFOULI et al., 2018; FAN et al., 2015; GUO et al., 2017a; GUO et al., 2016a; GUO et al., 2017b; HAN et al., 2018; HAWELKA et al., 2013; HUSAIN; DIH, 2012; LEAL et al., 2018;

LUZ et al., 2013; RONG, 2016; SHEN et al., 2016; XIONG et al., 2017; YIFAN et al., 2016; YOUNG; VERHULST, 2017; ZHANG et al., 2016a; ZHANG et al., 2016b). Em geral, os autores utilizam-se das fontes de Big Data para o desenvolvimento ou para a melhoria nos resultados e no desempenho de plataformas que: compilam, qualificam e disponibilizam informações e ofertas turísticas; ou descrevem e sugerem atrações, rotas e pontos de interesse.

5 CONCLUSÃO

O presente estudo corrobora a relevância do tema associado à possibilidade de inovação trazida pela utilização de Big Data para a construção de conhecimento turístico, especialmente sob a forma de indicadores, a considerar a crescente e recente produção científica identificada.

Em termos quantitativos, a revisão apontou para uma produção predominantemente europeia, asiática e norte-americana. No que concerne às possíveis fontes de dados, aquelas ligadas à World Wide Web são predominantes, sendo o Twitter a plataforma mais utilizada. Das fontes não ligadas à WWW, os dados de empresas de telefonia são os mais utilizados.

Por fim, as principais aplicações práticas aqui encontradas se referem: à estimação ou à projeção da quantidade ou demanda de turistas de entrada ou saída; ou à compreensão da satisfação, opinião, sentimentos e interesses manifestos. Fora do objetivo central do estudo, relevantes revisões bibliográficas foram encontradas, assim como diversas aplicações de Big Data com finalidade de construção de sistemas turísticos de informação e recomendação.

BIG DATA FOR TOURISM INDICATORS GENERATION: DATA SOURCES AND APPLICATIONS

ABSTRACT

The goal is to summarize the scientific production on the use of Big Data for calculating tourism indicators regarding possible data sources, practical applications and main entities involved with research in this area. A bibliographic research has been conducted in the form of a systematic literature review on articles retrieved April 23, 2019 from Scopus and Web of Science databases. The analysis of 175 scientific articles revealed predominance of a European, Asian and American scientific production using sources related to the World Wide Web to estimate tourist's quantity and sentiment. Hopefully the results will be able to enable tourism knowledge producers and users to envision innovative and alternative, more cost-effective, methods of generating tourism indicators.

KEYWORDS: *Big Data.
Tourism. Indicators.
Knowledge. Innovation.*



REFERÊNCIAS

- ALIS, C. et al. *Estimating tourism statistics with Wikipedia page views*, p. 1-2, 2015.
- BAGGIO, R. Measuring Tourism: Methods, Indicators, and Needs. In: FAYOS-SOLÀ, E.; COOPER, C. (Eds.). *The Future of Tourism*. Cham: Springer, 2019.
- BERZINA, I.; LAUBERTE, I. *The model of automation and extension of tourism economic impact assessment in specific regions*, p. 195-202, 2018.
- BIN, C. et al. Personalized POIs Travel Route Recommendation System Based on Tourism Big Data. In: PACIFIC RIM INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 15., Nanjing, 2018. *Proceedings...* Nanjing: 2018.
- BOULAALAM, O. et al. Proposal of a Big data System Based on the Recommendation and Profiling Techniques for an Intelligent Management of Moroccan Tourism. *Procedia Computer Science*, v. 134, p. 346-351, 2018.
- BROWN, S. L.; EISENHARDT, K. M. Product development – Past research, present findings, and future-directions. *Academy of Management Review*, v. 20, n. 2, p. 343-378, 1995.
- CENTOBELLI, P.; NDOU, V. Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism*, p. 1-22, 2019.
- CITRO, C. F. From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*, v. 40, p. 137-161, 2014.
- CHEN, C.; CHEN, H. *Construction of Local Tourist Information Platform Based on Cloud Computing*, 2017.
- CLASTER, W.; GHOTBI, N.; SHANMUGANATHAN, S. *Gathering Medical Tourism Information through Algorithmic Text Analysis of Tweets*, 2015.
- DEMUNTER, C. Tourism Statistics: Early Adopters Of Big Data? In: UNWTO INTERNATIONAL CONFERENCE ON TOURISM STATISTICS, 6., [s.l.]. *Proceedings...* [s.l.]: UNWTO, 2017.
- DEZFOULI, M. B.; SHAHRAKI, M. H.; ZAMANI, H. A Novel Tour Planning Model using Big Data. 2018 *International Conference on Artificial Intelligence and Data Processing*, p. 1-6, 2018.
- DONOVAN, C.; FLAHERTY, E. T.; HEALY, E. Q. Using big data from Wikipedia page views for official tourism statistics. *Statistical Journal of the IAOS*, (2017), v. 33, p. 1-7, 2017.
- DRUCKER, P. F. *Sociedade Pós-Capitalista*. São Paulo: Pioneira, 1993.
- EUROSTAT. *Scheveningen Memorandum on Big Data and Official Statistics*, 2013. Disponível em: <<http://ec.europa.eu/eurostat/documents/42577/43315/Scheveningen-memorandum27-09-13>>. Acesso em: 15 out. 2019.
- EUROSTAT. *Methodological manual for tourism statistics*. Luxemburgo: Office for the Official Publications of the European Communities, 2014.
- FAN, R.; DU, J.; ZHOU, Y. *Study and Implementation of Accurate Retrieval System Based on Attractions Interest Model*. v. 336, p. 477-485, 2015.
- FENG, N.; LI, J.; ZHANG, G. *A Study of the Structure of China's Mainstream Online Tourism Information Network Based on SNA*. 2013.
- FORBES, D. A. Strategies for Managing the Behavioural Symptomatology Associated with Dementia of the Alzheimer Type: A Systematic Overview. *Canadian Journal of Nursing Research Archive*, v. 30, n. 2, 1998.
- GARCÍA, F. C. et al. Quality implications of the use of big data in tourism statistics: Three exploratory examples. *Boletín de Estadística e Investigación Operativa*, v. 34, p. 138-148, 2018.
- GIL, A. C. *Métodos e Técnicas de Pesquisa Social*. São Paulo, Atlas, 2008.
- GREEN, S.; HIGGINS J. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5. 2005.
- GUARDA, T. et al. *Pervasive Smart Destinations*. 2018.
- GUO, T. et al. *CrowdTravel: Leveraging Heterogeneous Crowdsourced Data for Scenic Spot Profiling and Recommendation*. v. 9917, p. 617-628, 2016a.

- GUO, Y., BARNES, S., JIA, Q. Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation. *Tourism Management*, v. 59, p. 467-483, 2016b.
- GUO, T. et al. CrowdTravel: scenic spot profiling by using heterogeneous crowdsourced data. *Journal of Ambient Intelligence and Humanized Computing*, v. 9, 2017a.
- GUO, T. et al. CrowdTravel: leveraging heterogeneous crowdsourced data for scenic spot profiling. *Journal of Zhejiang University (Engineering Science)*, v. 51, n. 4, p. 663-668, 2017b.
- HACKL, P. Big Data: What Can Official Statistics expect? *Statistical Journal of the IAOS*, v. 32, p. 43-52, 2016.
- HAN, X. et al. Tourism Information Push System Based on Convolutional Neural Network. *E3S Web of Conferences*. 53. 2018.
- HAWELKA, B. et al. Geo-Located Twitter as Proxy for Global Mobility Patterns. *Cartography and Geographic Information Science*, v. 41, 2013.
- HELBING, D. et al. *Behavioural Control or Digital Democracy? - A Digital Manifesto*. 2016.
- HU, F. et al. A graph-based approach to detecting the tourist movement patterns using social media data. *Cartography and Geographic Information Science*, 2018.
- HUSAIN, W.; DIH, L. Y. A Framework of a Personalized Location-based Traveler Recommendation System in Mobile Application. *International Journal of Multimedia and Ubiquitous Engineering*, v. 7, 2012.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Código de Boas Práticas das Estatísticas do IBGE*. Rio de Janeiro: IBGE, 2013.
- ICHIFUJI, Y. et al. *A study for understanding of tourist person trip pattern based on log data of Wi-Fi access points*, p. 2167-2174, 2016.
- JANNUZZI, P. de M.; CRACIOSO, L. de S. Produção e disseminação da Informação Estatística pelas Agências Estaduais no Brasil. *Revista São Paulo em Perspectiva*, v. 16, n.3, 2002.
- KITAYAMA, D. *Extraction method for Anaba spots based on name recognition and user's evaluation*. iiWAS. 2016.
- KITCHIN, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Londres: Sage, 2014.
- KITCHIN, R. Big Data and Official Statistics: Opportunities, Challenges and Risks. *Statistical Journal of the International Association of Official Statistics*, v. 31, n. 3, p. 471-481, 2015.
- KOIDE, N. et al. *Estimation of national tourism statistics based on Wi-Fi association log data*, p. 2175-2179, 2016.
- LEAL, F. et al. *Analysis and Visualisation of Crowd-sourced Tourism Data*, p. 98-101, 2016.
- LEAL, F. et al. *Semantic Profiling and Destination Recommendation based on Crowd-sourced Tourist Reviews*, p. 140-147, 2018.
- LI, J. et al. Big data in tourism research: A literature review. *Tourism Management*, v. 68, p. 301-323, 2018.
- LIANG, S., Schuckert, M., Law, R. & Guo, X. (2017). A Retrospective Analysis and Future Trends of E-Tourism Research in China. *Journal of China Tourism Research*. 13.
- LIU, Y.; SHI, J. How inter-city high-speed rail influences tourism arrivals: evidence from social media check-in data. *Current Issues in Tourism*, v. 22, n. 9, p. 1025-1042, 2019.
- LUZ, N. et al. Collective intelligence in toursplan: An online tourism social network with planning and recommendation services. *ACM International Conference Proceeding Series*, p. 42-48, 2013.
- LIU, P. et al. Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. *Tourism Management*, v. 73, p. 105-114, 2019.

- MA, Y. et al. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *International Journal of Hospitality Management*, v. 71, p. 120-131, 2018.
- MARIANI, M. et al. Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 2018.
- MAYER-SCHONBERGER, V.; CUKIER, K. *Big data: A Revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, 2013.
- MENCHAVEZ, J. C.; ESPINOSA, K. J. P. *Fun in the Philippines: Automatic Identification and Sentiment Analysis of Tourism-Related Tweets*, p. 660-667, 2015.
- MILLER, H. J. The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, v. 50, p. 181-201, 2010.
- MOHER, D. et al. Reprint - preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Physical therapy*, v. 89, n. 9, p. 873-880, 2009.
- NOZAWA, Y. et al. *Inferring Tourist Behavior and Purposes of a Twitter User*, p. 101-112, 2017.
- ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÓMICO (OCDE). Rumo a um Ambiente Sustentável: Indicadores Ambientais. In: OCDE. *Série Cadernos de Referência Ambiental*. v. 9. Salvador: OCDE, 2002.
- ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÓMICO (OCDE). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. 3. ed. Paris: OECD Publishing, 2005.
- OLIVEIRA, R. A. de; BARACHO, R. M. A. The development of tourism indicators through the use of social media data: the case of Minas Gerais. *Brazil Information Research*, v. 23, n. 4, 2018.
- ONDER, I., KOERBITZ, W.; HUBMANN-HAIDVOGEL, A. Tracing Tourists by Their Digital Footprints: The Case of Austria. *Journal of Travel Research*, v. 55, 2014.
- PACHECO, R. C. D. S. Dados e Governo Abertos na Sociedade do Conhecimento. *Linked Open Data – Brasil*. Florianópolis-SC, 2014.
- PACHECO, R. C. D. S. Coprodução em Ciência, Tecnologia e Inovação: Fundamentos e Visões. In: PEDRO, J. M.; FREIRE, P. de S. (Orgs.). *Interdisciplinaridade: Universidade e Inovação Social e Tecnológica*. Curitiba: CRV, 2016.
- PARK, E. et al. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Current Issues in Tourism*, p. 1-7, 2018.
- PLENTZ, N. D.; BERNARDES, M. M. e S.; FRAGA, P. G. R. *Sistema de indicadores de inovação, competitividade e Design para empresas desenvolvedoras de produtos: versão 01*. Porto Alegre: Marcavisual, 2015.
- PRAMANA, S. et al. *Big data for government policy: Potential implementations of bigdata for official statistics in Indonesia*, p. 17-21, 2017.
- PROVENZANO, D.; HAWELKA, B.; BAGGIO, R. The mobility network of European tourists: a longitudinal study and a comparison with geo-located Twitter data. *Tourism Review*, v. 73, n. 1, p. 28-43, 2018.
- RAI, A. Evaluating the image of healthcare hubs online - A cross sectional analysis of google and face book reviews of private hospitals of Kolkata, eastern India. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XLII, n. 5, p. 893-899, 2018.
- RONG, F. *Design of Tourism Resources Management Based on Artificial Intelligence*, p. 436-439, 2016.
- SCHMIDT, E.; COHEN, J. *A nova era digital: como será o futuro das pessoas, das nações e dos negócios*. Rio de Janeiro: Intrínseca, 2013.

- SCHNORR-BAECKER, S. Statistical monitoring systems to inform policy decision-making, and new data sources. *Statistical Journal of the IAOS*, v. 33, p. 1-12, 2016.
- SCHREIBER, A. Th. et al. *Knowledge engineering and management: the CommonKADS methodology*. Cambridge: MIT Press, 2000.
- SHEN, J. et al. Landmark Reranking for Smart Travel Guide Systems by Combining and Analyzing Diverse Media. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 46, p. 1-13, 2016.
- SHI, B.; ZHAO, J.; CHEN, Po-Ju. Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data. *Current Issues in Tourism*, v. 20, n. 11, p. 1186-1209, 2017.
- SHIMADA, K. et al. *Analyzing Tourism Information on Twitter for a Local City*, 2011.
- SHOVAL, N.; AHAS, R. The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, v. 18, p. 1-20, 2016.
- TIDD, J.; BESSANT, J.; PAVITT, K. *Gestão da inovação*. Tradução Elizamari Rodrigues Becker et al. 3. ed. Porto Alegre: Bookman, 2008.
- THELWALL, M. *Sentiment Analysis for Tourism: Managerial Approaches, Techniques, and Applications*, 2019.
- SETIADI, Y.; ULUWIYAH, A. *Improving data quality through big data: Case study on big data-mobile positioning data in Indonesia tourism statistics*, p. 43-48, 2017.
- UNITED NATIONS. *International Recommendations for Tourism Statistics 2008 (online)*, 2010. Disponível em: <<http://unstats.un.org/unsd/tradeserv/tourism/manual.html>>. Acesso em: 15 out. 2019.
- UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE). *Classification of Types of Big Data*, 2013. Disponível em: <<https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>>. Acesso em: 15 out. 2019.
- UNITED NATIONS STATISTICAL COMMISSION (UNSC). Big Data and Modernisation of Statistical Systems. In: SESSION OF THE STATISTICAL COMMISSION, 45., 2014, Nova York. *Proceedings...* Nova York: 2014. Disponível em: <<http://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf>>. Acesso em: 15 out. 2019.
- WORLD TOURISM ORGANISATION (UNWTO). *Sustainable tourism indicators and destination management: regional workshop*. Madri: UNWTO, 2007.
- WORLD TOURISM ORGANISATION (UNWTO). *Methodological Notes to the Tourism Statistics Database*. Madri: UNWTO, 2017.
- VENZIN, M.; KROGH, G. von; ROOS, J. Future Research into Knowledge Management. KROGH, G. von et al. (Eds.). *Knowing in Firms - Understanding, Managing and Measuring Knowledge*. Londres: SAGE, 1998.
- XIONG, H. et al. *A novel recommendation algorithm frame for tourist Spots based on multi-clustering bipartite graphs*, p. 276-282, 2017.
- YANG, K. et al. Urban tourism research based on the social media check-in data. In: INTERNATIONAL CONFERENCE ON SMART AND SUSTAINABLE CITY (ICSSC 2017), 4., 2017, Xangai. *Proceedings...* Xangai: ICSSC, 2017.
- YIFAN, Y. et al. Design and implementation of tourism activity recognition and discovery system. In: WORLD CONGRESS ON INTELLIGENT CONTROL AND AUTOMATION (WCICA), 12., 2016, Guilin. *Proceedings...* Guilin: WCICA, 2016.
- YOUNG, A.; VERHULST, S. *Jamaica's interactive community mapping: Open data and crowdsourcing for tourism*, 2017.
- ZHANG, Z.; PAN, H.; XU, G. Research on personalized tourism attractions recommendation model based on user social influence. *Advances in Modelling and Analysis B*, v. 59, p. 62-75, 2016a.
- ZHANG, Z. et al. A Context-Awareness Personalized Tourist Attraction Recommendation Algorithm. *Cybernetics and Information Technologies*, v. 16, 2016b.

SOBRE OS AUTORES



Alexandre Augusto Biz

Professor do Departamento de Engenharia do Conhecimento – UFSC. Professor Colaborador do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – UFSC.

Pós-Doutorado Empresarial CNPq realizado no Instituto Stela. Doutor em Engenharia e Gestão do Conhecimento (UFSC). Mestre em Turismo e Hotelaria (UNIVALI). Graduado em Turismo e Hotelaria (UNIVALI). É líder do Grupo de Pesquisa TURITEC. Coordenador do Laboratório de Destinos Turísticos Inteligentes. Desenvolve pesquisas em Destinos Turísticos Inteligentes, Aplicações da Engenharia e Gestão do Conhecimento para o Turismo. Revisor de periódicos nacionais e internacionais. Coordenador Geral do Seminário Brasil Espanha de Inovação Tecnológica em Turismo <www.destinosinteligentes.com.br>.



João Artur Souza

Professor titular da Universidade Federal de Santa Catarina no Departamento de Engenharia do Conhecimento. Graduação em Matemática (Licenciatura) pela Universidade Federal de

Santa Catarina (1989) e em Direito pela Universidade do Sul Catarinense. Mestrado em Matemática e Computação Científica pela Universidade Federal de Santa Catarina (1993). Doutorado em Engenharia de Produção pela Universidade Federal de Santa Catarina (1999). Trabalhou na Universidade Federal de Pelotas de 1993 a 2007 como professor na área de Matemática, atuando especialmente em Educação a Distância. Foi coordenador dos Cursos de Graduação em Matemática e Matemática a Distância (2005-2006). Pesquisa na área de Inovação, Inteligência Artificial, Gestão do

Conhecimento, Gestão de Risco e Controle Interno e Universidade Corporativa e faz parte de três grupos de pesquisa IGTI: Inteligência, Gestão e Tecnologias para Inovação (IGTI) (Líder), Engenharia da Integração e Governança do Conhecimento (ENGIN) e Interdisciplinar em Conhecimento, Aprendizagem e Memória Organizacional (KLON). Ministra as disciplinas Projetos Interdisciplinares, Gestão de Inovação e Lógica Matemática na Graduação. Na pós-graduação, atua no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da UFSC na área de Inteligência e Gestão para Inovação.

Lattes: <<http://lattes.cnpq.br/6695591100082194>>.



Rafael Bassegio Caumo

Estatístico pela UFRGS, mestre pela ENCE/IBGE e doutorando no EGC/UFSC. Possui longo currículo de atuação com pesquisas quantitativas e análise de dados, sendo atualmente pro-

fessor de Métodos Estatísticos e Análise Estatística na ESAG/UEDESC e consultor estatístico autônomo. Possui conhecimento técnico sobre modelos estatísticos e de machine learning, amostragem e inferência, com aplicações de interesse nas áreas de Big Data para construção de indicadores socioeconômicos. Já foi por duas vezes presidente do Conselho Regional de Estatística da 4ª Região e passou por instituições como Fundação de Economia e Estatística (chefe de centro de pesquisa), Pesquisa de Emprego e Desemprego da Região Metropolitana de Porto Alegre (estatístico responsável), Secretaria da Saúde do RS, entre outras.
