

BIG DATA COMO FUENTE DE CONOCIMIENTO TURÍSTICO. ESPECIAL REFERENCIA AL OPEN DATA Y AL BIG DATA SOCIAL

Alfonso Cerezo Medina *, Antonio Guevara Plaza ** y Manuel Enciso García-Oliveros ***

Resumen: El objetivo de este artículo es resaltar el valor estratégico que posee el Big Data para el sector turístico. El creciente volumen de información generada a nivel global, derivada de la huella digital debe verse como una oportunidad para adaptarse al nuevo perfil del turista, más dinámico, exigente y conectado que busca experiencias integrales y personalizadas. El Big Data se presenta como un recurso que puede dar respuesta a este reto, pero debemos conocer en qué consiste, dónde se encuentra y cómo ponerlo en valor. Para ello, en este trabajo se realiza una aproximación al concepto de Big Data, se presentan las principales fuentes de documentación turística y se dan a conocer algunos casos de éxito de Big Data Social aplicado al turismo. Finalmente se plantean los principales beneficios que puede aportar el Big Data al turismo y los riesgos inherentes.

Palabras clave: Big Data, Big Social Data, Open Data, datos, fuentes de datos, Turismo.

Abstract: The aim of this paper is to highlight the strategic role of Big Data for the tourism sector. The growing volume of information created all over the world as a consequence of the tourist's digital footprint, should be seen as an opportunity to better satisfy more dynamic and complex tourist profiles. Hence, this research explores the concept of Big Data, presenting its main sources for tourism and showing some study cases of success where Big Social Data has been applied. Finally, the main benefits and challenges of Big Data and tourism are considered.

Key words: Big Data, Big Social Data, Open Data, data sources, tourism.

I. INTRODUCCIÓN

Al igual que la tecnología, el turismo es una fuerza globalizadora que ha contribuido al crecimiento que se viene produciendo en los últimos años en el volumen de información generada, y es que el turístico, no es precisamente un sector ajeno a la digitalización experimentada por la sociedad en las últimas décadas, hasta el punto de que puede afirmarse que es uno de sus máximos exponentes. Esto se explica porque el turismo es intensivo en información (Sheldon,

1997) dada la naturaleza de los servicios turísticos (Koo, Gretzel, Hunter, & Chung, 2015; McCole, 2002; Sheldon, 1997; Werthner & Klein, 1999). Este hecho junto a diversas cifras como el volumen de viajeros internacionales, los nuevos hábitos de los turistas y la aparición de recursos turísticos más flexibles y colaborativos, han configurado un escenario ideal para el auge del Big Data (BD) en el sector.

Las últimas cifras de la Organización Mundial del Turismo (OMT) señalan que

* Facultad de Turismo. Universidad de Málaga.

** Departamento de Lenguajes y Ciencias de la Computación. Universidad de Málaga.

*** Departamento de Lenguajes y Ciencias de la Computación. Universidad de Málaga.

en 2017 se produjeron más de 1.322 millones de viajes internacionales. A esto debe añadirse que el turista actual está cada vez más hiperconectado y es multicanal, gracias a la aplicación y uso de las tecnologías de la información y la comunicación (TIC), destacando en los últimos cinco años el auge de los dispositivos móviles a la hora de conectarse en Internet, (el 59% de los de las personas entre 16 y 74 años de la Unión Europea en 2016) (1). Son numerosos los datos que retratan a un turista digital como, por ejemplo, que los sectores con mayores ingresos por comercio electrónico sean las agencias de viajes, operadores turísticos y el transporte aéreo (2), que el 95% de los viajeros utilicen recursos digitales a lo largo de todo el ciclo de consumo turístico (3), que los sitios web que recogen y muestran opiniones y valoraciones de otros viajeros (Ewom), junto a los medios sociales, sean tras las recomendaciones de amigos y familiares, las fuentes de información más importantes a la hora de planificar un viaje (4), que en España el sector de los viajes, transporte y turismo sea el segundo más seguido en redes sociales (5) o que un tercio de los españoles tenga instalada alguna aplicación (app) de viajes y turismo (6). Estamos por tanto, ante un nuevo tipo de turista más exigente que ha desarrollado nuevas costumbres relacionadas con el uso de las tecnologías (Caro Herrero, Luque, & Zayas, 2015), que dejan una huella digital antes, durante y después del viaje; desde la planificación del mismo, pasando por la reserva y el pago, hasta el consumo y la posterior acción de compartir la experiencia turística en los medios sociales (Fuchs, Höpken, & Lexhagen, 2014).

El escenario descrito anteriormente sugiere porqué en el sector turístico existe una ingente cantidad de datos, que, si son extraídos, curados y analizados de forma conveniente, pueden generar información relevante para la toma de decisiones, tanto en el ámbito empresarial como en la gestión de destinos turísticos. Aquí es donde entra en juego el Big Data (BD), sobre cuya adopción y aplicación basada en infraestructuras y servicios TIC (social media, sensores, dispositivos móviles...) recae el concepto de turismo inteligente (Gretzel, Sigala, Xiang, & Koo, 2015), y determina la capacidad de creación de conocimiento, y por tanto la competitividad de los destinos (Shaw & Williams, 2009).

La OMT es consciente del papel crucial que puede jugar el BD a la hora de implementar políticas y acciones turísticas de una forma integrada y sostenible, siendo prueba de ello la declaración de Muscat (7), en la que se hace hincapié sobre la importancia de generar mejor información para el turismo cultural a partir del uso de las fuentes de datos existentes y el aprovechamiento del BD para medir y visualizar sinergias entre el turismo y la cultura. Por esta razón, la institución internacional está desarrollando en la actualidad el estudio sobre BD y Turismo Cultural, que se centra en cómo el BD está siendo aplicado en museos y sitios Patrimonio de la Humanidad en la planificación, gestión y comercialización del turismo cultural en distintos destinos.

En definitiva, en un contexto globalizado, digital y más competitivo que nunca, la implementación de las técnicas analíticas

del BD se justifica como herramienta para dar respuesta a las nuevas necesidades de los turistas, permitiendo ofrecer una mejor experiencia turística (Zhu et al., 2014), creación de ofertas contextualizadas, co-creación de productos y servicios y proveer valor en tiempo real (Neuhofner, Buhalis, & Ladkin, 2012). En el momento actual, el siguiente paso que debe dar el sector turístico es aprovechar su privilegiada posición en cuanto a la cantidad y calidad de los datos de que se dispone, para seguir siendo un sector pionero en el aprovechamiento de las herramientas y aplicación de las TICs. A pesar del reconocimiento de las oportunidades que ofrece el BD, la investigación sobre cómo los destinos turísticos pueden aprovechar los grandes volúmenes de datos generados por los turistas para la creación de valor es escasa y se precisa de un análisis más profundo (Vecchio, Mele, Ndou, & Secundo, 2017).

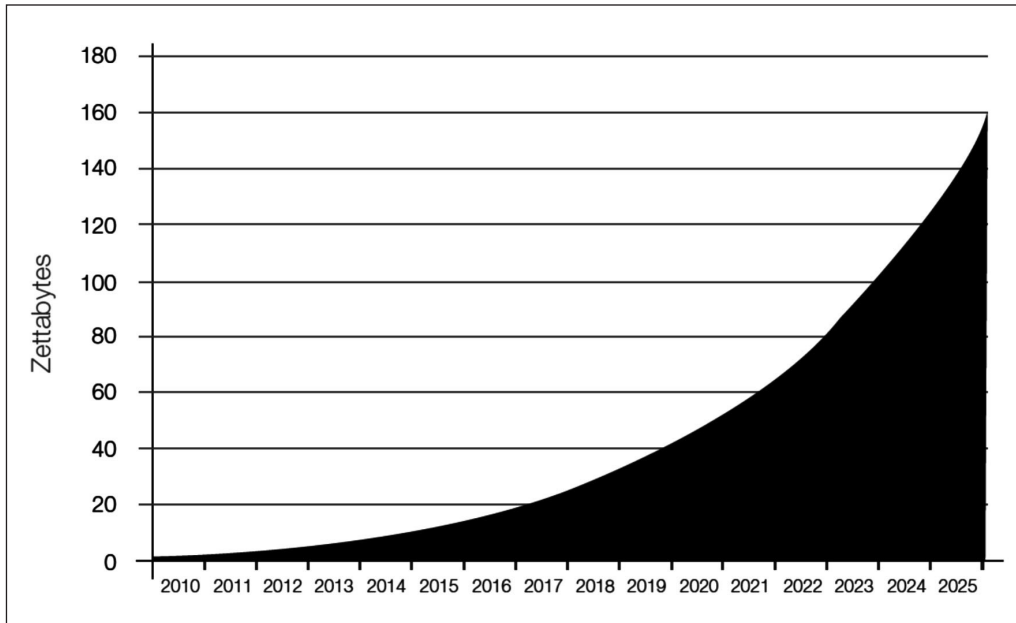
Una vez presentada las sinergias que se dan entre turismo y BD, la estructura del artículo quedaría del siguiente modo: en la sección 2 se introduce una aproximación al concepto Big Data y su popularización en los últimos años. En la siguiente sección se realiza una descripción de los recursos accesibles de información, las principales fuentes de datos y documentación turística, así como de las herramientas tecnológicas susceptibles de convertirlas en conocimiento. Seguidamente, en la sección 4, se enumeran algunos casos de éxito de BD social aplicados al turismo. Por último, se concluye el artículo considerando una serie de retos y riesgos asociados al BD.

II. APROXIMACIÓN A LA DEFINICIÓN DE BIG DATA

En los últimos años se viene asistiendo a un crecimiento continuado del volumen de datos generados, tal y como se aprecia en el gráfico 1, estimándose que en la próxima década la cantidad de datos global se multiplique por diez (Reinsel, Gantz, & Rydning, 2017). En 2016 se generaron 16 zettabytes, lo que equivaldría a grabar una película HD de cientos de millones de años de duración. Este incesante auge se explica por factores como los avances realizados en la capacidad de computación de los sistemas informáticos, en la capacidad de almacenamiento de los mismos y en su abaratamiento, a lo que se une el desarrollo de Internet y la mejora continua en los sistemas de telecomunicaciones. Todo ello ha hecho posible un crecimiento sin precedentes y que en la actualidad pueda decirse que nos encontremos en la era de los datos o revolución de los datos, donde todo ciudadano y dispositivo deja un rastro digital en el ciberespacio que representa una oportunidad para las empresas y las instituciones, que pueden convertir dichos datos en información útil si cuentan con las herramientas y mecanismos adecuados de extracción, procesamiento y visualización.

No cabe duda de que el término Big Data (BD) ya forma parte de la jerga habitual de los negocios, y pese a que existen referencias que indican que se comenzó a hablar de este fenómeno a mediados de los 90 (Gandomi & Haider, 2015), no es hasta 2011 cuando se popularizó, como así queda reflejado en el gráfico 2, en el que se repre-

Gráfico 1
Evolución de la generación de datos a escala global



FUENTE: IDC's Data Age 2025 study (2017)

Gráfico 2
Evolución de la popularidad de Big Data en las búsquedas de Google a nivel global



FUENTE: Google trends.

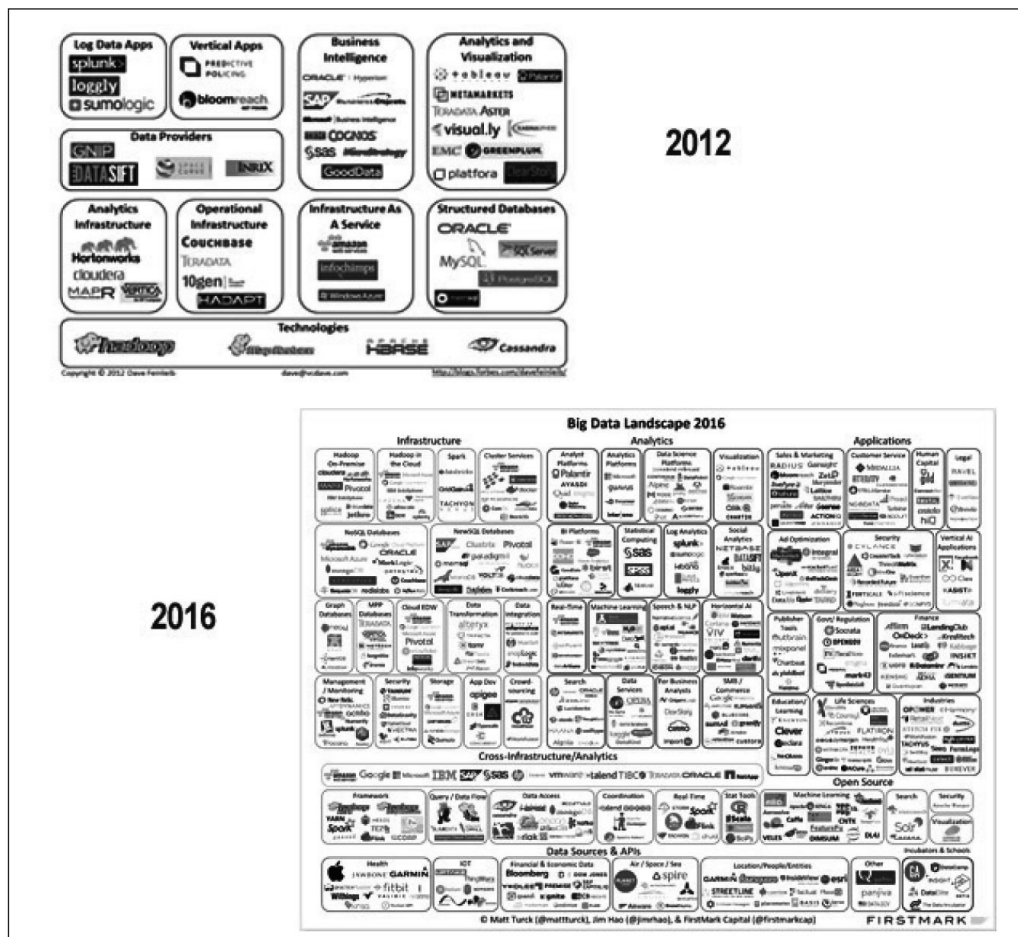
senta la evolución creciente del interés de búsqueda de Big Data hasta la actualidad.

Otro aspecto que ayuda a comprender el vertiginoso auge del BD en el mundo en los últimos años, es la evolución del panorama (véase Figura 1), que permite visualizar de una forma rápida el crecimiento significativo que se ha producido en torno al BD,

tanto en en el número como en los tipos (infraestructuras, aplicaciones, fuentes, empresas, etc.)

Existe cierta confusión sobre el significado del BD, y probablemente existan más artículos tratando de definir el concepto de Big Data que aquellos que realmente lo aplican (Demunter, 2017). En efecto, no existe

Figura 1
Evolución del panorama del BD 2012-2016



FUENTE: Elaboración propia a partir de Dave Feinleib 2012 y Matt Truck, Jim Hao y FirstMark Capital 2016.

una acepción única del concepto (Sánchez, 2016), ya que es cambiante y ha sido definido desde distintas esferas como la académica y la empresarial. A partir de una revisión de la literatura, se ha podido observar que la gran mayoría de definiciones se pueden dividir en dos grandes grupos:

1. las que hacen referencia a un volumen de datos
2. las que lo hacen a tecnologías (herramientas informáticas o infraestructuras) y técnicas.

Entre las que conciben el BD como un volumen de datos, ítem 1 anterior, no se identifica una frontera mínima en cuanto al tamaño, que varía constantemente y puede oscilar desde terabytes hasta zettabytes, aunque la gran mayoría sí coinciden en resaltar que se tratan de volúmenes de datos cuyo tamaño supera la capacidad de almacenamiento, procesamiento, análisis y visualización de las herramientas y hardware de uso común.

Del proceso de revisión de definiciones también es posible extraer que la gran mayoría de autores comparten la misma visión a la hora de definir las características o dimensiones que deben poseer los volúmenes de datos, las llamadas “V”: velocidad (de procesamiento y análisis), variedad (tipos de datos estructurados, semiestructurados y no estructurados provenientes de diversas fuentes como vídeo, texto, audio, sensores, medios sociales, webs, etc.), volumen, valor (conversión de los datos en información de calidad) y veracidad (fiabilidad de los datos).

Del mismo modo, existe un consenso generalizado sobre la finalidad que posee el BD, que no es otra que la de generar valor a partir de una mejor comprensión de los datos y por ende una toma de decisiones más eficiente, dicho de otro modo, solucionar problemas reales de forma optimizada, reduciendo costes, en tiempo real, obteniendo ingresos adicionales y anticipando riesgos.

III. TIPOLOGÍA DE DATOS Y FUENTES

Cuando se hace referencia a la variedad de los datos, se alude a la estructura de los mismos, pudiendo distinguir entre datos estructurados, semiestructurados y no estructurados.

- Los datos estructurados son aquellos que presentan un esquema rígido y bien definido, como los que pueden tabularse en filas y columnas. Tradicionalmente son aquellos datos alfanuméricos que se pueden encontrar en una base de datos relacional o en una hoja de cálculo, permitiendo la realización de consultas precisas y un procesamiento automatizado.
- Los datos semiestructurados están a medio camino entre los estructurados y no estructurados. Son aquellos que presentan una cierta estructura pero que es irregular; contienen etiquetas o marcadores que les permiten ser leídos por máquinas. El lenguaje XML (Extensible Markup Language) es un ejemplo típico de dato semiestructurado, así como los weblogs. Estos datos presentan capacidades de consulta limitadas dada la escasa rigidez de su esquema.

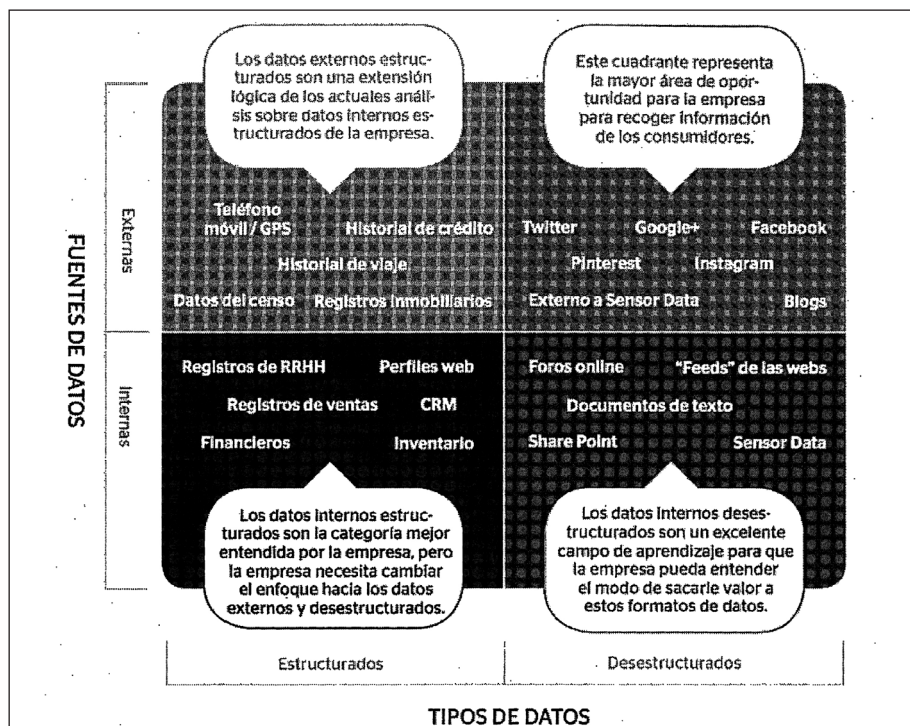
- Los datos no estructurados son aquellos que no poseen una estructura organizacional uniforme, no presentan ningún esquema más allá de la mera secuencia de bytes. Texto (publicaciones en redes sociales, mensajes de whatsapp, etc.), imágenes, audio y vídeo constituyen los ejemplos más comunes de este tipo de datos. IDC (International Data Corporation) estima que el 90% de BD son datos no estructurados.

Según la fuente de origen de los datos, estos pueden ser internos o externos. Si fuésemos una empresa, los datos internos serían

los derivados de nuestra actividad, como registros de ventas, recursos humanos, información de nuestros clientes, inventario, etc. Los externos son datos generados por otros actores que pueden representar una oportunidad para la creación de conocimiento, tales como estadísticas, censos, redes sociales, redes de telefonía, pagos con tarjetas, etc. La figura 2 representa un claro esquema de los datos digitales que pueden generarse en la actualidad en el sector turístico en función de la fuente y el tipo.

En los últimos años ha aparecido un fenómeno que está afectando a las fuentes y

Figura 2
Tipos de datos y fuentes de origen



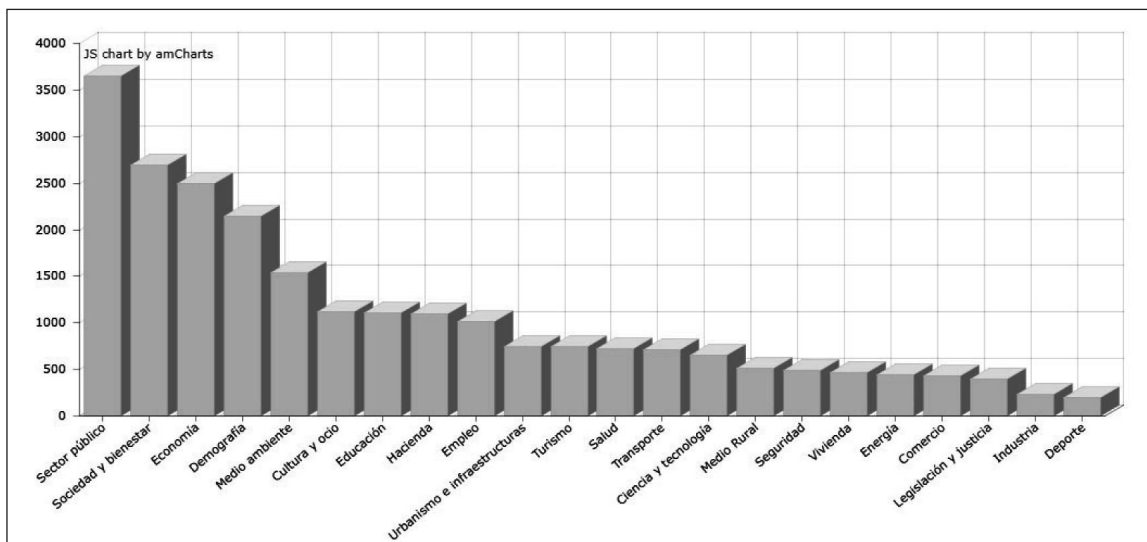
FUENTE: Sánchez, 2016.

tipos de datos descritas. Los datos abiertos (OD) u “open data” son aquellos que las instituciones públicas comparten con la doble finalidad de alcanzar una mayor transparencia y permitir la reutilización de información pública (Bauzá, 2017). Es un movimiento que se está extendiendo a nivel internacional con el apoyo de dichas instituciones públicas y que cuenta con políticas y marcos normativos que lo regulan (Sevillano, 2015), siendo por tanto una parte importante de la Agenda Digital de la Unión Europea y otras instituciones; se ha convertido en un factor importante en el intercambio de conocimientos (López, Serna, Gerrikoitia, & Murua, 2013). La información que dispone la administración pública derivada del ejercicio de sus funciones incluye muchos tipos de datos, que pueden posibilitar la monitorización y ajuste de las polí-

ticas públicas y mejorar la calidad de vida de los ciudadanos a través del desarrollo de aplicaciones móviles (Sevillano, 2015), pudiendo también contribuir al crecimiento económico y la generación de riqueza y empleo (Bauzá, 2017). Estas afirmaciones pueden sustentarse en casos de éxito como los de Madiva o Bloowatch. La primera es una startup que a partir de datos abiertos del catastro y de las ofertas publicadas en portales inmobiliarios, permite conocer el precio aproximado de compraventa de una vivienda. La segunda, utiliza los datos abiertos de meteorología para generar planes de navegación y planificación de actividades marítimas para escuelas náuticas y empresas de turismo activo.

En la actualidad existen decenas de países que cuentan con plataformas de datos

Gráfico 3
Número de conjuntos de datos publicados por cada categoría (8)

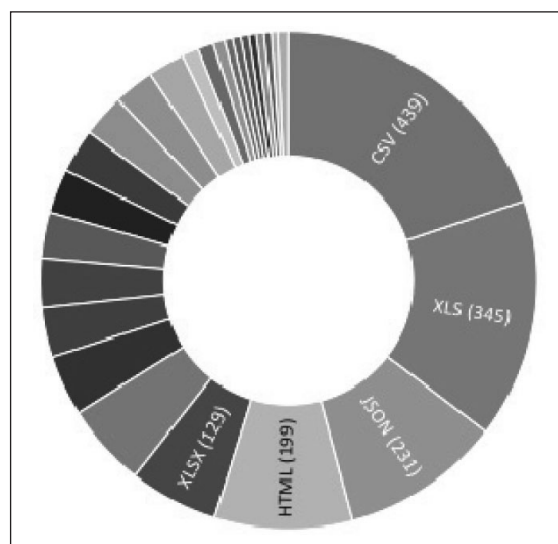


FUENTE: datos.gob.es

abiertos, entre las que se encuentra **Aporta**, una iniciativa promovida por el Ministerio de Energía, Turismo y Agenda Digital, que se lanzó en 2009 con el fin de promocionar la cultura de la apertura de información en España, creando las condiciones para el desarrollo del mercado de la reutilización de la información del sector público. Esta iniciativa ofrece 16.544 conjuntos de datos, de los cuales 747 son relativos al caso de uso de turismo, que previamente han sido procesados y transformados, e incluyen distintos aspectos de la oferta y la demanda turística como índices de precios, número de plazas de alojamiento, número de empresas, movimientos de viajeros, encuestas de gasto turístico, recursos turísticos, etc. El gráfico 3 permite ver que el turismo ocupa un puesto intermedio en cuanto a volumen de datos en comparación con otras categorías.

Atendiendo al formato en el que los conjuntos de datos se encuentran disponibles, en el gráfico 4 se observa que el más frecuente en la categoría turismo es CSV, seguido de

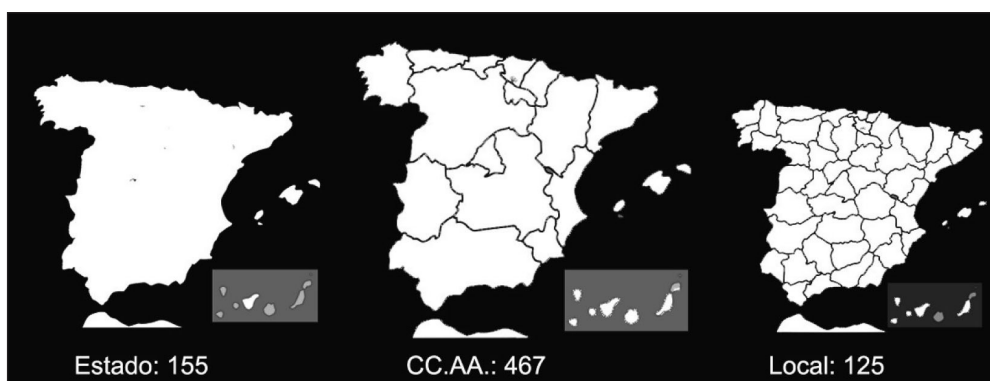
Gráfico 4
Conjunto de datos de turismo según formato



FUENTE: Elaboración propia a partir de datos.gob.es

Xls y Json, representando entre los tres casi el 50% del total. En total se identifican has-

Figura 3
Conjuntos de datos en turismo publicados según Administración



FUENTE: Elaboración propia a partir de datos.gob.es

ta 31 tipos de formatos (9) diferentes, entre los que destacan un 10% que corresponden a formatos de geolocalización.

Otro aspecto que Aporta permite conocer es el publicador de los datos abiertos, lo que facilita crear rankings según la categoría de datos. Según la administración pública responsable de la generación de los datos, las Comunidades Autónomas son con diferencia las que más conjuntos de datos han publicado en turismo.

Finalmente, a través de la plataforma de datos abiertos del gobierno de España se puede conocer quiénes son las instituciones que más publican dentro de cada categoría. Así, en turismo puede apreciarse en el cuadro 1 que la Comunidad Autónoma del País

Vasco y el Instituto Nacional de Estadística son los organismos que más datos abiertos han publicado.

III.1. Fuentes de datos

Las fuentes y tipos de datos citados hasta el momento pueden recoger información de los recursos turísticos del destino y la huella digital de los visitantes en forma de opiniones y tracking espacio-temporal, lo que conforma la piedra angular de la gestión de los destinos basada en el conocimiento (López et al., 2013).

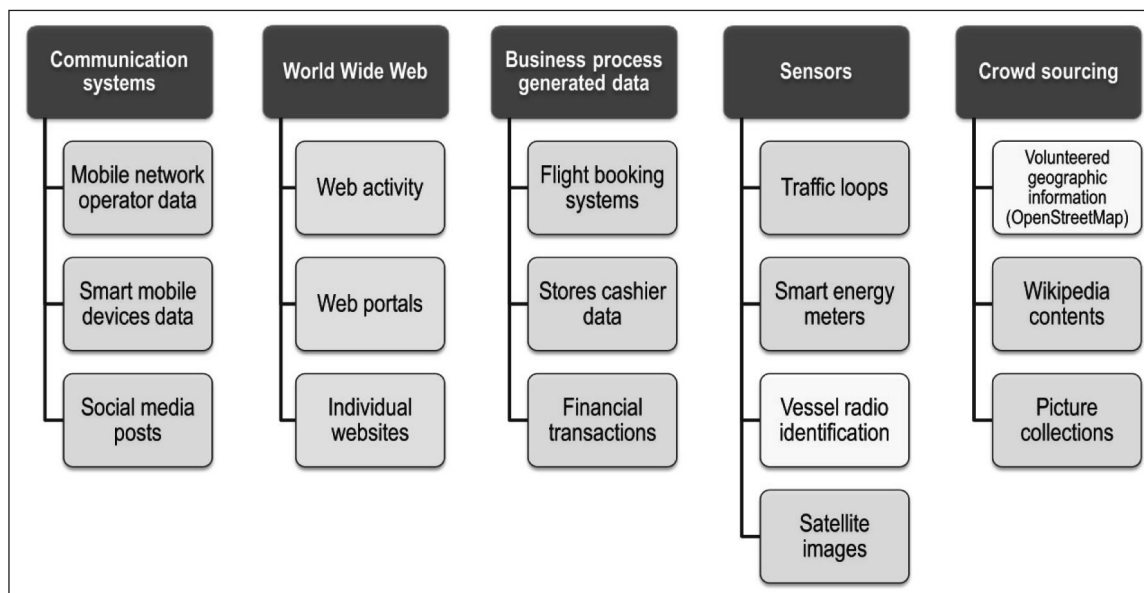
Según Demunter (2017), en la figura 4 se esbozan las fuentes de BD más comunes en turismo, señalando que muchas de ellas

Cuadro 1
Ranking de publicadores de OD en turismo

Comunidad Autónoma de País Vasco (318)	Ayuntamiento de Torrent (13)
Instituto Nacional de Estadística (140)	Ayuntamiento de Las Palmas de Gran Canaria (12)
Gobierno de Aragón (75)	Ayuntamiento de Valencia (10)
Ayuntamiento de Gijón (31)	Ayuntamiento de Madrid (10)
Comunidad Autónoma de Castilla-La Mancha (30)	Diputación Provincial de Cádiz (8)
Generalitat Valenciana (23)	Diputacion Provincial de Barcelona (7)
Junta de Castilla y León (16)	Ayuntamiento de Zaragoza (6)
Comunidad Autónoma de País Vasco (318)	Ayuntamiento de Santander (6)
Instituto Nacional de Estadística (140)	Instituto de Turismo de España (5)
Gobierno de Aragón (75)	Cabildo Insular de Tenerife (5)
Ayuntamiento de Gijón (31)	Ayuntamiento de Cáceres (5)
Comunidad Autónoma de Castilla-La Mancha (30)	Ministerio de Energía, Turismo y Agenda Digital (4)
Generalitat Valenciana (23)	Ayuntamiento de Lorca (4)
Junta de Castilla y León (16)	

FUENTE: datos.gob.es

Figura 4
Fuentes de BD para el turismo



FUENTE: Demunter, 2017

no resultan nuevas para la estadísticas oficiales, y otras llevan siendo utilizadas desde hace bastante tiempo.

En el esquema de fuentes de BD propuesto por Demunter queda patente que una parte significativa de las mismas la constituyen sensores tecnológicos, (torres de telefonía, smartphones, cajeros automáticos, cámaras de vídeo, sensores de tráfico, imágenes satélite, sitios web, etc.), no obstante, otras fuentes están adquiriendo un protagonismo cada vez mayor. Es el caso de la actividad del turista en los medios sociales, lo que podría calificarse como una especie de “sensor humano”; que se considera una de las principales fuentes de generación de BD en turismo y ofrece enormes posibilidades para una toma de decisiones más eficiente

(Miah, Vu, Gammack, & McGrath, 2017), tanto, que ha dado lugar al concepto de “Big Data Social”, que puede definirse como los “procesos y métodos que son diseñados para proporcionar a cualquier empresa o usuario un conocimiento relevante y sensible a partir de las fuentes de datos de los medios sociales, que pueden caracterizarse por sus diferentes formatos y contenidos, su gran tamaño y la generación de la información en línea” (Bello-Orgaz, Jung, & Camacho, 2016).

El Big Data Social (BDS) es el resultado del explosivo crecimiento que ha tenido en la última década el contenido generado por los usuarios (*User Generated Content*, UGC en sus siglas en inglés) gracias a Internet y al uso de las tecnologías móviles que

hacen posible que el turista pueda dejar una valoración de una empresa o servicio turístico, documentarse para realizar un viaje, o subir una foto o vídeo a una red social. El UGC ha evolucionado desde los años 90 e incluye comunidades virtuales, plataformas de opiniones de consumidores, blogs, plataformas de microblogging, redes sociales y wikis. Esta nueva fuente de BD no está pasando desapercibida para el sector turístico, y en los últimos años se han desarrollado desde el ámbito académico distintas investigaciones que representan diferentes aplicaciones del BD en la vida real, tales como medir la imagen de un destino, clasificar los establecimientos de alojamiento en categorías distintas a las clásicas estrellas o mejorar la competitividad y resultados financieros de empresas. En un apartado posterior se especificarán más detalles de tales investigaciones.

III.2. Fuentes de documentación turística

Hasta el momento se han presentado distintos esquemas de fuentes turísticas de BD, destacando por su volumen y posibilidades de reutilización las siguientes: datos estadísticos (INE y Eurostat), datos generados por sensores tecnológicos, el Big Data Social y los datos abiertos, sin embargo, aún pueden identificarse otras fuentes que presentan una gran potencialidad para la generación de conocimiento en turismo. Nos referimos a los catálogos de producciones científicas, repositorios y bases de datos de documentación turística, que pueden proporcionar datos estructurados o semiestructurados en formato xml, bib, ris o txt (10) sobre distintas entidades como artículos,

libros, revistas, proyectos, obras audiovisuales...

Los datos contenidos en estas fuentes de documentación turística son susceptibles de ser curados, procesados y analizados mediante técnicas de BD que ofrezcan como resultados estudios bibliométricos y análisis de redes sociales que sirvan para determinar por ejemplo qué se investiga en materia turística en un determinado espacio geográfico-temporal, quiénes llevan a cabo tales investigaciones o cuáles son los temas que más interés despiertan en la comunidad académica, en definitiva, se podría generar un conocimiento con el fin de detectar campos de estudio sobresaturados y nuevos nichos, un aspecto especialmente relevante en turismo, dónde no existe un área de conocimiento que sirva como guía para la investigación en la que se especifiquen cuáles son los temas prioritarios para la sociedad (Peláez & Ferrera, 2017).

A continuación se presentan las fuentes de documentación turísticas más importantes a nivel nacional e internacional.

- *Hospitality and tourism complete (EBSCO).*

Es una base de datos que cubre todas las áreas del campo de la hostelería y el turismo, incluyendo textos completos de revistas académicas, revistas, informes, libros y periódicos. Este catálogo incluye más de 500 publicaciones en modo texto completo y casi 1,4 millones de registros que hacen referencia a documentos desde 1924 hasta la actualidad. Esta base de datos recopila los registros de tres

conocidas colecciones: la base de datos Hospitality de Cornell University, Articles in Hospitality and Tourism (AHT), previamente a cargo de las universidades de Surrey y Oxford Brookes y Lodging, Restaurant & Tourism Index (LRTI), anteriormente elaborada por Purdue University.

- *Sistema de información científica del turismo (SICTUR).*

Esta fuente compila y pone a disposición de los profesionales y académicos toda la producción científica e investigadora en turismo realizada por las universidades españolas en los últimos 15 años. Es una base de datos que incluye la producción científica de más de 2.500 investigadores, lo que se traduce en más 4.000 artículos de revistas, más de 1000 actas de congresos, más de 3.500 comunicaciones a congresos, más de 2.000 libros y más de 1.100 proyectos de investigación y transferencia.

- *Biblioteca virtual OMT eLibrary.*

Esta biblioteca virtual ofrece acceso a un amplio número de publicaciones de la OMT (Organización Mundial del Turismo) en sus respectivos idiomas (hay ya 650 libros en español, francés, inglés, ruso o árabe). Incluye información estadística en el repertorio de datos de turismo Tourism Statistics, ordenada por países. La biblioteca virtual de la OMT es una fuente de información que pone a disposición de los Estados Miembros, Miembros Afiliados y suscriptores una cantidad inigualable de estudios e in-

formación sobre turismo. Las materias abarcan todos los temas importantes que impulsan al sector turístico e incluyen, entre otros, el ecoturismo, el desarrollo sostenible, las finanzas y la inversión, la gestión de riesgos y situaciones de crisis, los estudios de mercado, las estadísticas del turismo y la lucha contra la pobreza. La biblioteca virtual de la OMT contiene libros, publicaciones periódicas e informes estadísticos por país y por indicador.

- *Centro de Documentación Turística de España (CDTE).*

El CDTE está considerado uno de los centros de documentación sobre turismo más importante del mundo, tanto por el número de documentos que custodia, como por la calidad de los mismos. Creado en 1985, es un servicio de Turespaña que recopila, analiza, custodia y difunde la información y los documentos relativos al turismo, ofrece al usuario más de 120.000 referencias documentales informatizadas referidas a carteles y folletos turísticos, revistas y monografías especializadas en turismo, publicaciones institucionales y legislación.

IV. APLICACIONES DERIVADAS DEL SOCIAL BIG DATA EN TURISMO

De forma paralela al importante crecimiento seguido por el UGC en medios sociales, han proliferado en turismo las investigaciones que utilizan este tipo de contenidos como una alternativa a la investigación mediante encuestas. De hecho, se han

desarrollado a lo largo de la última década investigaciones que se enmarcan en el ámbito del Social Big Data y el turismo que constituyen ejemplos de cómo la gestión de datos puede traducirse en conocimiento para asegurar una mejor toma de decisiones. Así queda recogido en el estudio de revisión bibliográfica realizado por Lu y Stepchenkova (2015), que analizaron una muestra de más de 120 artículos científicos escritos en inglés y publicados en revistas y congresos desde 2001 hasta abril de 2013. Los autores identificaron los principales temas turísticos abordados en este campo, los métodos aplicados al UGC, y el software usado para extraer y analizar el UGC. Asimismo, los autores pudieron identificar los tipos de da-

tos UGC empleados, como textos, fotos, vídeos y valoraciones de usuarios, siendo los predominantes los primeros. A partir de este trabajo, el cuadro 2 muestra la distribución de las investigaciones de BD social en turismo en función del subsector en el que se realizó, las fuentes UGC y los temas objeto de estudio. El 83% de los estudios se enmarcaron en el área de los destinos turísticos y el alojamiento; el tipo de fuente predominante fueron las plataformas de opiniones de usuarios (Tripadvisor), empleadas en el 43% de los casos. Finalmente, respecto a los temas abordados por las investigaciones, cabe señalar que en el 40% de los casos analizados, la calidad del servicio fue el tema central de análisis, seguido por la imagen de

Cuadro 2
Distribución de investigaciones sobre UGC en turismo

SUBSECTORES	Destinos (54)
	Alojamiento (48)
	Restauración (8)
	Transporte aéreo (4)
	Curceros (2)
FUENTES DE UGC	Plataformas de opiniones de usuarios (63)
	Blogs (33)
	Sitios para compartir vídeos (20)
	Redes sociales (11)
	Comunidades virtuales (8)
	Wikis (1)
TEMAS ABORDADOS	Calidad del servicio (50)
	Imagen de destino y reputación (26)
	eWom (25)
	Comportamiento del consumidor (14)
	Patrones de movilidad (7)

FUENTE: Elaboración propia partir de Lu y Stepchenkova (2015).

destino y el boca-oído electrónico, con un 20% respectivamente.

A partir de 2013 han seguido publicándose investigaciones en el campo de estudio del BDS y el turismo, siendo posible encontrar en España distintos estudios de caso, que en su gran mayoría han tratado de analizar la imagen online de los destinos turísticos a partir del UGC en medios sociales como blogs y sitios web de reseñas de viajes. En este sentido destaca el estudio de Chen, Guevara y Alarcón (2017), que proponen un método para estudiar las opiniones de los turistas chinos sobre el destino España a partir de comentarios de Facebook, lo que permite determinar qué le interesa a los turistas chinos y cómo se forman sus opiniones. En una línea similar ha trabajado intensamente Mariné (2016, 2017a, 2017b; Mariné y Anton, 2015), cuyos estudios se han centrado en resaltar la utilidad del BD para los destinos turísticos inteligentes y estudiar la imagen online de Barcelona o París a través de los medios sociales, proponiendo para ello un marco teórico y métodos para analizar y medir la imagen de los destinos; los resultados obtenidos poseen implicaciones directas para los gestores de las políticas públicas de los destinos a la hora de evaluar sus estrategias de comunicación y marketing turístico.

Aparte de los estudios centrados en el análisis de la imagen de los destinos turísticos, destaca la investigación de Gemár y Jiménez (2015), que llevaron a cabo un estudio novedoso aplicando una herramienta de minería de texto para analizar los principales medios sociales, incluyendo Twitter, Facebook, LinkedIn, YouTube y otros, en

una muestra de hoteles. Las dimensiones analizadas fueron sentimientos, pasión y alcance, encontrando los autores una dependencia entre varias variables obtenidas a través de la minería de textos y el desempeño financiero. Los hallazgos de este trabajo indican, por tanto, que el análisis de los medios sociales utilizando estas técnicas puede ser un método para mejorar el rendimiento financiero.

El comportamiento espacial de los turistas también es objeto de investigación, y a él han contribuido recientemente Salas, Moya, Garcia y Gutierrez (2018), que a diferencia de gran parte de los estudios previos centrados en el uso de una sola fuente de datos, han indagado en el uso de varias para rastrear la huella digital del turista en el destino a través de sus actividades. A partir de Panooramio (visitas turísticas), Foursquare (consumo) y Twitter, caracterizan los espacios turísticos de Madrid, detectando así zonas multifuncionales y zonas especializadas.

Otro caso práctico de BDS y turismo a resaltar por su carácter innovador es la reciente investigación de Martín, Fernández, Mateu y Mariné, (2018), que han creado un modelo informático que a través de machine learning es capaz de clasificar a los establecimientos hoteleros y viviendas de uso turístico a partir de 18 millones de comentarios de clientes en todo el mundo. Los resultados de esta investigación pueden ayudar los turistas en su toma de decisiones y puede también resolver cuestiones como la estandarización de los criterios utilizados en los diferentes países para categorizar los alojamientos. En la misma línea Martín (2016) realizó un estudio previo sobre los sistemas

de clasificación hotelera a nivel mundial. A partir de una muestra de más de 14.000 hoteles en 100 ciudades de todo el mundo extraída de Booking y Tripadvisor, comprobó si el sistema de clasificación por estrellas, el precio de las habitaciones e incluso el tamaño de los hoteles coincidían con la satisfacción del usuario desde el punto de vista de las puntuaciones realizadas en las citadas plataformas.

El fenómeno del BDS no se limita de forma exclusiva a las publicaciones en revistas científicas o congresos, pues se trata de un tema de gran actualidad y relevancia social, territorial y económica sobre el que se están llevando a cabo proyectos concebidos como mecanismos de respuesta a los retos de la economía y sociedad digital incluidos en el marco del Plan Estatal de Investigación Científica y Técnica de Innovación, como el desarrollo, innovación y adopción de soluciones y tecnologías ligadas al open/linked/Big Data y la reutilización de la información del sector público generando valor y conocimiento, o los medios sociales como vehículos de prestación de servicios y potencial desarrollo empresarial. Entre ellos destaca el proyecto “Análisis turístico de las plataformas de alojamiento colaborativo en destinos españoles a través de contenido generado por los usuarios y otras fuentes en línea (TURCOLAB)”, en el que participan más de una decena de investigadores de la Universidad de Lleida y que aborda el estudio de las plataformas de comercialización de alojamientos de turismo colaborativo desde el punto de vista de la oferta y la demanda, a partir del análisis turístico del contenido generado por los usuarios y otras fuentes en línea. TURCOLAB se plantea como ob-

jetivo científico desarrollar un modelo analítico, conceptual y metodológico, integral e innovador para comprender el fenómeno del alojamiento de turismo colaborativo, identificar las ventajas que aporta a sus usuarios, evaluar su presencia en Internet, interactividad e impacto social, tanto para sus usuarios como para el sector de alojamiento y turístico general, que sea de utilidad para el sector turístico, investigadores y gestores públicos. Los principales aportes del proyecto serán el desarrollo, creación y aplicación de una nueva metodología para el análisis de grandes datos en plataformas turísticas colaborativas y, a partir de los resultados obtenidos, la construcción de un nuevo marco conceptual para entender las plataformas turísticas colaborativas que vayan más allá de su mera ecuación con los medios sociales y demuestre su valor añadido frente a otros tipos de alojamiento

En último lugar se presenta Smart Data Andalucía, un ejemplo de transferencia de conocimiento que representa una solución práctica y real de la aplicación del BDS al turismo. Se trata de una herramienta tecnológica puesta en marcha en Andalucía que permitirá a destinos y empresas conocer mejor al cliente al que se dirigen a partir del análisis cuantitativo y cualitativo de las opiniones de los viajeros monitorizadas en 2.000 medios distintos. Es un proyecto transversal desarrollado por la administración pública en la que participan la Consejería de Turismo y Deporte de la Junta de Andalucía, el Centro de Innovación Turística Andalucía Lab, el Sistema de Análisis y Estadística del Turismo de Andalucía (SAETA) y la Universidad de Málaga.

V. DISCUSIÓN Y RETOS ACERCA DE BD Y TURISMO

A lo largo del presente artículo se ha tratado de exponer los principales tipos de datos y fuentes que pueden englobarse bajo el paraguas del Big Data y el turismo, prestando especial atención a las fuentes de documentación turística existentes, a los datos abiertos y al UGC. En este sentido se ha hecho referencia al papel estratégico que desempeñan actualmente los medios sociales como fuente de BD, en la que el contenido generado por el usuario se erige como uno de los máximos exponentes en el ámbito turístico. Seguidamente se han presentado diversas aplicaciones prácticas de BD al turismo en distintas áreas como la empresa y los destinos turísticos, tanto desde la academia como desde las instituciones públicas, mostrando de este modo los beneficios que el BD puede tener para el sector turístico.

A pesar de todo lo que se ha avanzado en los últimos años en este campo de investigación, es necesario incidir en una serie de retos que deben afrontarse todavía en materia de privacidad, seguridad, formación y visualización. Aún estamos ante una industria poco madura, que no puede ofrecer soluciones milagro, y es que a pesar de que el BD no haya dejado de ganar protagonismo, aún no puede sustituir otras fuentes de información como puedan ser las encuestas.

Independientemente de todos los beneficios que pueda aportar el BD al turismo, el punto de partida crucial es la detección de los casos de usos, es decir, plantear las preguntas adecuadas para poder solventar problemas reales, ya que unos de los prin-

cipales retos es encontrar los enfoques adecuados para analizar grandes cantidades de datos y además en tiempo real, pues de ello dependerá los métodos de clasificación y algoritmos a aplicar (Bello et al., 2016), muchos de ellos de muy reciente creación, cuyo conocimiento y dominio no es trivial (Ortega, 2017). Por tanto, la formación resulta imprescindible para que el manejo del BD pueda generar conocimiento turístico, siendo necesario crear e impulsar itinerarios formativos que desarrollen competencias digitales (matemáticas, estadística, programación, etc) que puedan ofrecer como resultado personal cualificado.

Desde el punto de vista de los tipos de datos y su generación desde los medios sociales, el UGC sigue evolucionando, aumentando su volumen y variedad de formatos, lo que requiere del desarrollo de nuevas técnicas y tecnologías de BD con mayor capacidad de extracción, almacenamiento, análisis e interpretación. A partir del análisis de la literatura y de los casos presentados, se ha comprobado que el análisis texto en el marco del UGC ha sido el más frecuente, sin embargo, existe una necesidad creciente de procesar e interpretar la fotografías e imágenes, así como investigar y comprender cómo interactúan texto, imágenes y otros elementos periféricos, para poder obtener una visión holística de porqué ciertos contenidos generados por los usuarios son más informativos y valiosos que otros (Ma, Xiang, Du, & Fan, 2018).

En materia de protección de datos y privacidad se observa que la realidad diverge de un ordenamiento jurídico que debe adaptarse de forma irremediable para garantizar

los derechos de los titulares de los datos, permitiendo el acceso y la reutilización de forma segura.

En definitiva, el auténtico valor de la relación entre BD y fuentes de datos y documentación turística radica en la capacidad de poder generar información que pueda ser visualizada y convertida en conocimiento, lo que a su vez podrá traducirse en una toma de decisión más eficiente en términos de sostenibilidad.

BIBLIOGRAFÍA

- BAUZÁ MARTORELL, F. J. (2017). Big data y open data en la administración turística: acceso y reutilización de información. *Revista Vasca de Administración Pública. Herri-Arduralaritzako Euskal Aldizkaria*, (108), 19–41.
- BELLO-ORGAZ, G., JUNG, J. J., & CAMACHO, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28, 45–59. <https://doi.org/10.1016/J.INFFUS.2015.08.005>
- CARO HERRERO, J. L., LUQUE, A., & ZAYAS, B. (2015). Nuevas tecnologías para la interpretación y promoción de los recursos turísticos culturales. *Pasos. Revista de Turismo Y Patrimonio Cultural*, 13(4), 931–945.
- CHEN, F.-W., GUEVARA PLAZA, A., & ALARCÓN URBISTONDO, P. (2017). Automatically extracting tourism-related opinion from Chinese social media. *Current Issues in Tourism*, 20(10), 1070–1087. <https://doi.org/10.1080/13683500.2015.1132196>
- DEMUNTER, C. (2017). Tourism statistics: early adopters of big data? In *Sixth UNWTO International Conference on Tourism Statistics MEASURING SUSTAINABLE TOURISM* (pp. 21–24).
- 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. <https://doi.org/10.1016/J.JDMM.2014.08.002>
- GANDOMI, A., & HAIDER, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/J.IJINFOMGT.2014.10.007>
- GEMÁR, G., & JIMÉNEZ-QUINTERO, J. A. (2015). Text mining social media for competitive analysis. *Tourism & Management Studies*, 11(1), 84–90.
- GRETZEL, U., SIGALA, M., XIANG, Z., & KOO, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179–188. <https://doi.org/10.1007/s12525-015-0196-8>
- KOO, C., GRETZEL, U., HUNTER, W. C., & CHUNG, N. (2015). Editorial : The Role of IT in Tourism. *Asia Pacific Journal of Information Systems*, 25(1), 99–104. <https://doi.org/10.14329/apjis.2015.25.1.099>
- LÓPEZ DE IPIÑA, D., SERNA NOCEDAL, A., GERRIKAGOITIA ARRIEN, J. K., & MURUA, I. (2013). Open Data en Turismo. *tourGUNE Journal of Tourism and Human Mobility*, 1, 1–10.
- LU, W., & STEPCHENKOVA, S. (2015). User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *Journal of Hospitality Marketing & Management*, 24(2), 119–154. <https://doi.org/10.1080/19368623.2014.907758>
- MA, Y., XIANG, Z., DU, Q., & FAN, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *International Journal of Hospitality Management*, 71, 120–131. <https://doi.org/10.1016/J.IJHM.2017.12.008>
- MARINE-ROIG, E. (2016). Online Travel Reviews: A Massive Paratextual Analysis. In *Analytics in Smart Tourism Design* (pp. 179–202). Springer, Cham. https://doi.org/10.1007/978-3-319-44263-1_11

- MARINE-ROIG, E. (2017). Measuring Destination Image through Travel Reviews in Search Engines. *Sustainability*, 9(8), 1425. <https://doi.org/10.3390/su9081425>
- MARINE-ROIG, E., & ANTON CLAVÉ, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162–172. <https://doi.org/10.1016/j.jdmm.2015.06.004>
- MARINE-ROIG, E., & ESTELA. (2017). Measuring Destination Image through Travel Reviews in Search Engines. *Sustainability 2017, Vol. 9, Page 1425*, 9(8), 1425. <https://doi.org/10.3390/SU9081425>
- MARTIN-FUENTES, E. (2016). Are guests of the same opinion as the hotel star-rate classification system? *Journal of Hospitality and Tourism Management*, 29, 126–134. <https://doi.org/10.1016/j.jhtm.2016.06.006>
- MARTIN-FUENTES, E., FERNANDEZ, C., MATEU, C., & MARINE-ROIG, E. (2018). Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management*, 69, 75–83. <https://doi.org/10.1016/J.IJHM.2017.10.016>
- MCCOLE, P. (2002). The role of trust for electronic commerce in services. *International Journal of Contemporary Hospitality Management*, 14(2), 81–87. <https://doi.org/10.1108/09596110210419264>
- MIAH, S. J., VU, H. Q., GAMMACK, J., & MCGRATH, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771–785. <https://doi.org/10.1016/j.im.2016.11.011>
- NEUHOFFER, B., BUHALIS, D., & LADKIN, A. (2012). Conceptualising technology enhanced destination experiences. *Journal of Destination Marketing & Management*, 1(1–2), 36–46. <https://doi.org/10.1016/j.jdmm.2012.08.001>
- ORTEGA FERNÁNDEZ, C. (2017). Cómo las empresas pueden impulsar su negocio a través de las plataformas e-commerce con el Big Data , el aprendizaje automático y el management científico. *Economía Industrial*, (405), 75–86.
- PELÁEZ, A., & FERRERA, M. (2017). The usefulness of social media analysis within scholarly publications: a study of first - tier tourism journals. *Tourism & Management Studies*, 13(1), 43–50.
- REINSEL, D., GANTZ, J., & RYDNING, J. (2017). *Data Age 2025: Don't Focus on Big Data; Focus on the Data That's Big Data Age 2025*:
- SALAS-OLMEDO, M. H., MOYA-GÓMEZ, B., GARCIA-PALOMARES, J. C., & GUTIERREZ, J. (2018). Tourists' digital footprint in cities: comparing big data sources. *Tourism Management*, 66, 13–25.
- SÁNCHEZ OCAÑA, P. (2016). Big data y turismo deportivo: un mundo de nuevas oportunidades. In *Turismo y deporte* (1st ed., pp. 33–60). Dykinson.
- SEVILLANO PÉREZ, F. (2015). Big Data. *Economía Industrial*, (395), 71–86.
- SHAW, G., & WILLIAMS, A. (2009). Knowledge transfer and management in tourism organisations: An emerging research agenda. *Tourism Management*, 30(3), 325–335. <https://doi.org/10.1016/J.TOURMAN.2008.02.023>
- SHELDON, P. J. (1997). *Tourism information technology*. CAB INTERNATIONAL.
- VECCHIO, P. DEL, MELE, G., NDOU, V., & SECUNDO, G. (2017). Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing & Management*. <https://doi.org/10.1016/j.ipm.2017.10.006>
- WERTHNER, H., & KLEIN, S. (1999). *Information Technology and Tourism: A Challenging Relationship by H Werthner, S Klein, Stefan Klein - New, Rare & Second-hand Books Online at Waterstones marketplace*.
- ZHU, H., CHEN, E., XIONG, H., YU, K., CAO, H., & TIAN, J. (2014). Mining Mobile User Preferences for Personalized Context-Aware Recommendation. *ACM Transactions on Intelligent Systems and Technology*, 5(4), 1–27. <https://doi.org/10.1145/2532515>

NOTAS

(1) Estadísticas sobre sociedad y economía digital - Hogares y particulares. EUROSTAT.

(2) CNMCDData. Comisión Nacional de los Mercados y la Comptencia. 2017.

(3) La Transformación digital en el sector turístico. Fundación Orange.

(4) Flash Eurobarometer 432. Preferences of Europeans toward torusim. 2016. European Comission.

(5) Estudio anual redes sociales 2017. IAB

(6) Estudio Anual Mobile Marketing 2017. IAB

(7) Muscat Declaration on Tourism and Culture: Fosterin Sustainable Development.

(8) Un mismo conjunto de datos puede pertenecer a varias categorías.

(9) CSV, XLS, JSON, HTML, XLSX, PC-Axis, KML, XML, XML-APP, API, RDF-XML, GeoJSON, RSS, RDF-Turtle, RDF-N3, Atom, JSON-LD, PDF, WMS, SHP, KMZ, OCTET-STREAM, GML, WFS, SPARQL, XHTML, Solr, ODS, GeoRSS, WMS-XML, TSV.

(10) Pueden proporcionar etiquetas como autores, títulos, palabras claves, fecha de publicación, etc.