

**MÁSTER UNIVERSITARIO EN ESTUDIOS AVANZADOS
EN DIRECCIÓN DE EMPRESAS**

BIG DATA Y PRIVACIDAD. ESTUDIO BIBLIOMÉTRICO

TRABAJO FIN DE MÁSTER





Departamento de Economía Financiera y Dirección de Operaciones

BIG DATA Y PRIVACIDAD. ESTUDIO BIBLIOMÉTRICO

Trabajo Fin de Máster presentado para optar al Título de Máster Universitario de Estudios Avanzados en Dirección de Empresas por Alejandro Muñoz Díaz, siendo los tutores del mismo los profesores Miguel Ángel Ríos Martín y Mariano Aguayo Camacho.

Vº. Bº. de los Tutores:

Alumno/a:

D/Dª. Miguel Ángel Ríos Martín

D/Dª. Alejandro Muñoz Díaz

D/Dª. Mariano Aguayo Camacho

Sevilla, 4 de Septiembre de 2017



**MÁSTER UNIVERSITARIO DE ESTUDIOS AVANZADOS EN
DIRECCIÓN DE EMPRESAS
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES**

**TRABAJO FIN DE MÁSTER
CURSO ACADÉMICO [2016-2017]**

TÍTULO:

BIG DATA Y PRIVACIDAD. ESTUDIO BIBLIOMÉTRICO

AUTOR/A:

Alejandro Muñoz Diaz

TUTOR/A:

Profesores Miguel Ángel Ríos Martín y Mariano Aguayo Camacho

LÍNEA DE TRABAJO:

Sistemas de Información

RESUMEN:

Revisión bibliográfica sobre la Privacidad de los datos personales en la actividad relacionada con el concepto "Big Data".

PALABRAS CLAVE:

Big Data, Data privacy, Privacy, Data protection, Data mining.

AGRADECIMIENTOS

Quiero manifestar la absoluta entrega que he tenido de mi madre y de mi padre, y con especial cariño de Ana, por el apoyo moral en los momentos más delicados, disposición, prioridad y la ayuda científica para perfilar y procurar la excelencia durante la realización de este proyecto de investigación.

También quiero reconocer la continua dedicación y el compromiso personal que han venido otorgándome a lo largo de todo el recorrido de este trabajo y su modo de infundir interés en la investigación de mis tutores Miguel Ángel Ríos Martín y Mariano Aguayo Camacho.

Mi gratitud a la Facultad de Ciencias Económicas y Empresariales de la Universidad de Sevilla, donde me he formado y que me ha permitido desarrollar este proyecto.

ÍNDICE

CAPÍTULO 1. INTRODUCCIÓN.....	9
1.1. Introducción.....	9
1.2. Justificación.....	11
1.3. Objetivos	13
1.4. Metodología	15
CAPÍTULO 2. BIG DATA Y PRIVACIDAD.....	17
2.1. Big Data	17
2.1.1. Concepto	17
2.1.2. Tipos de información manejados en Big Data	18
2.1.3. Características de los datos de Big Data	19
2.1.4. Ciclo de vida de los Datos	21
2.1.5. Herramientas de Big Data Analytics.....	22
2.2. La privacidad de la información en el Big Data	26
2.2.1. Concepto de Privacidad	26
2.2.2. El desafío de la seguridad y la Privacidad.	27
2.2.3. La seguridad en los métodos analíticos de Big Data	29
2.2.4. Posibles Técnicas para proteger la Privacidad	30
CAPÍTULO 3. ESTADO DE LA CUESTIÓN.....	35
CAPÍTULO 4. OTROS ESTUDIOS BIBLIOMÉTRICOS SOBRE BIG DATA Y TECNOLOGÍAS DE LA INFORMACIÓN. LA IMPORTANCIA DE ESTUDIOS ANÁLOGOS.....	43
CAPÍTULO 5. ANÁLISIS BIBLIOMÉTRICO	49
5.1. Metodología	49
5.2. Análisis de Datos.....	52
5.2.1. Número de artículos por año.....	52
5.2.2. Número de artículos por país	54
5.2.3. Número de artículos por revista	55
5.2.4. Número de artículos por universidad o entidad productora	60
5.2.5. Estudio de artículos por autores.....	66
5.2.6. Estudio de las Keywords encontradas en la revisión bibliográfica	67
CAPÍTULO 6. CONCLUSIONES.....	76
6.1. Epígrafe 1. Limitaciones del estudio:.....	79
6.2. Epígrafe 2. Futuras líneas de investigación:	80
CAPÍTULO 7. BIBLIOGRAFÍA.....	81

Relación de Figuras

Ilustración 1: Forma de trabajar de Hadoop	22
Ilustración 2: Ecosistema Hadoop.....	23
Ilustración 3: ¿Cómo trabaja MapReduce?	24
Ilustración 4: Interfaz de Cassandra aplicándose gracias al soporte de Aqua Studio..	26

Relación de Tablas

Tabla 1: Artículos por revista	55
Tabla 2: Relación de artículos por entidad productora	60
Tabla 3: Tabla de autores con dos artículos	66
Tabla 4: Keywords más comunes (o de primer nivel).....	67
Tabla 5: Keywords menos comunes (o de segundo nivel)	69
Tabla 6: Keywords de tercer nivel.....	69
Tabla 7: Keywords de interés futuro.....	71
Tabla 8: Keywords métodos de Privacidad	71
Tabla 9: Keywords de ámbito informático	72
Tabla 10: Keywords de ámbito gubernamental o legal.....	72
Tabla 11: Keywords de ámbito empresarial	73
Tabla 12: Keywords de ámbito ético o filosófico.....	73
Tabla 13: Keywords de ámbito Marketing y Comunicación	74
Tabla 14: Keywords de Riesgo u Oportunidad.....	74
Gráfica 1: Interés a lo largo del tiempo en el concepto.....	36
Gráfica 2: Interés por región	37
Gráfica 3: Keywords más vinculadas	37
Gráfica 4: Interés a lo largo del tiempo, “búsqueda web”	38
Gráfica 5: Interés por región en “búsquedas web”	39
Gráfica 6: Keywords más vinculadas a “búsquedas web”	39
Gráfica 7: Interés a lo largo del tiempo de “Big Data Privacy”	40
Gráfica 8: Interés a lo largo del tiempo de “Big Data Privacy”	40
Gráfica 9: Número de artículos por año	52
Gráfica 10: Número de artículos por país.....	54

CAPÍTULO 1. Introducción

1.1. Introducción

La tecnología ha avanzado enormemente en todos los ámbitos de la vida en las últimas décadas, convirtiendo los teléfonos inteligentes, portátiles y ordenadores en herramientas indispensables en el día a día.

El concepto que vamos a estudiar y analizar es conocido mediante multitud de nombres en todos los estudios e investigaciones manejados, los mayormente utilizados o los más comúnmente encontrados son “datos masivos”, “macro datos”, “inteligencia de datos” o “datos a gran escala”, pero por ser el mayormente utilizado y el término que más proliferado encontramos en la doctrina especializada en el tema, escogeremos Big Data para referirnos al mismo el resto del estudio.

Este concepto nació para darle una denominación al conjunto de datos en tamaño masivo, de tal forma que su volumen alcanza tal magnitud, que las aplicaciones informáticas tradicionales de procesamiento de datos no son lo suficientemente cualificadas o potentes para llevar a cabo su manipulación y análisis necesario por parte de las empresas.

Esta disciplina se enmarca dentro del sector “Tecnologías de la Información y la Comunicación” (mayormente conocido como TIC), ocupándose en general de las tareas de búsqueda, recolección, almacenamiento, visualización, análisis y compartición de los datos masivos que manejan. Las principales causas por las que se utilizan las herramientas de tratamiento de Big Data es para la creación de informes estadísticos o modelos predictivos en diferentes ámbitos, como pueden ser la publicidad, los datos de seguridad en la salud, los negocios o incluso la lucha contra las organizaciones terroristas.

En general, los altos cargos de las empresas opinan (tras tratar e implantar herramientas de Big Data en sus empresas), que los principales beneficios aportados son los siguientes y por este orden de preferencia (Big Data Soft, 2013):

- 1- Una nueva forma de acceder a la información interna y externa que maneja la empresa.
- 2- Información verídica y actualizada en tiempo real de la empresa y la competencia en el mercado.
- 3- Una forma de calcular la influencia de los datos en las nuevas tecnologías.

- 4- Una forma muy novedosa de obtener información de distintas y variadas fuentes.
- 5- Una herramienta muy potente con la que recabar información importante en los medios sociales y las redes sociales.

Se trata de una herramienta útil con bastante calado y gran potencial aun por explorar, en cuanto que su adaptabilidad es máxima y su usabilidad podrá llegar a ser tanto como la inventiva de las empresas que la utilicen (Pérez, 2015). Ahora bien, todo esto conlleva que exista una acumulación de datos importante, en su mayoría relativos a personas, y dicho depósito de datos debe ser manejado con una normativa determinada y específica, pues como podemos observar en la doctrina especializada, las normativas convencionales que han protegido los datos personales en Internet parece que han quedado obsoletas (Pérez, 2015).

Según la directiva 95/46/CE¹ se define al tratamiento de datos personales con esta clase de herramientas masivas como: “cualquier operación o conjunto de operaciones, efectuadas o no mediante procedimientos automatizados, y aplicadas a datos personales, como la recogida, registro, organización, conservación, elaboración o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma que facilite el acceso a los mismos, cotejo o interconexión, así como su bloqueo, supresión o destrucción”.

La protección de los datos íntimos en Big Data debe ofrecer una garantía suficiente que permita al individuo gestionar y delimitar las personas o entidades que puedan acceder a su información personal, es decir, que nadie nada más que el individuo objeto utilice sus datos relativos a su ambiente y realidad personal.

En la época actual, la sociedad tecnológica, el individuo ha asistido a la creación de una serie de circunstancias por las cuales el tratamiento de datos personales se hace obligatorio e inevitable si queremos mantener el tipo de vida (entretenimiento en directo, acceso a noticias, etc...) y servicios que son propios de la sociedad en la que vivimos y nos desarrollamos; de igual forma debemos ser conscientes de que si bien es obligatorio estar sometidos a este control de nuestros datos personales, de igual forma las empresas e instituciones deben estar sometidos a un entramado mínimo de criterios y normas que nos aporte un colchón de seguridad y evite que dicha información, en definitiva, pueda transformarse en una amenaza para la privacidad de las personas.

¹ Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (DO L 281, de 23 de noviembre de 1995).

Cuando convergen las inmensas posibilidades de la tecnología en la sociedad actual con la anteriormente comentada facilidad de manejar datos de las personas por parte de las empresas y organizaciones, se da la existencia de un posible estado de vulnerabilidad, cuyos damnificados, incluso, pueden llegar a desconocer su situación. En este apartado es clave el Big Data, de tal forma que hace unas décadas las posibilidades que la tecnología convencional aportaba en el manejo de los datos personales referidos a la intimidad eran mucho más limitadas, hoy día dichas posibilidades se multiplican exponencialmente, tanto por el potencial computacional como por las herramientas disponibles para depurar y tratar los datos (Kshetri, 2014).

El estado actual del derecho en cuanto a este tema, se encuentra en un estado obsoleto, puesto que no ofrece soluciones a las herramientas que la tecnología ha aportado en los últimos años. Actualmente las técnicas de extracción de datos son muy potentes y han experimentado una revolución, a la que normas como la mencionada anteriormente ya no son aplicables, por no adecuarse a la realidad, es por ello que la primera de las conclusiones alcanzadas es que el derecho aplicable en este sector debe actualizarse urgentemente para dar cabida a un nuevo tipo de protección a los derechos de la información íntima de la persona (Pérez, 2015).

Como se puede observar, este tema tiene mucho recorrido y denota un interés máximo en muchos investigadores y docentes (Bull-berg & Olsson, 2015), que han tratado el tema desde perspectivas y puntos de vistas muy diferentes, todos ellos válidos, y que en este trabajo se ha tratado de recopilar, extraer su información y obtener un punto de vista general del conocimiento alcanzado sobre la Privacidad manejada con herramientas de Big Data.

1.2. Justificación

La justificación de la realización de este proyecto académico es debida a la gran notoriedad que está teniendo esta tecnología en nuestros días. Cualquier persona con o sin conocimientos tecnológicos tiene cierto interés en la forma en que las empresas almacenan toda la información que se genera en el mundo virtual: en *Facebook*, *Twitter*, *Instagram* o cómo *Google* es capaz de manejar todas las transacciones que se hacen a diario, y sobre todo, qué grado de intimidad tiene el ciudadano frente al uso indebido que las empresas y los gobiernos les puedan dar a todos esos datos. *Big Data* alcanza todos los ámbitos: economía, educación, ciencias tecnológicas y naturales, lucha contra el crimen, etc...

La cantidad de datos que se genera actualmente es ingente, averiguar mediante qué herramientas se gestiona toda esa información y cómo dichas herramientas productivas tienen límites para no violar nuestra privacidad debe ser por lógica un motivo de interés para desarrollar no solo este, sino todos los estudios que se deba, tanto para conocer como para controlar el estado en el que nuestro derecho a la intimidad es protegido o violado.

La tecnología que estudiamos está íntimamente ligada con el sector empresarial, puesto que se conoce que la Administración Pública ya trabaja con este tipo de herramientas, son las empresas del sector privado sobre las que más control se debe aplicar para controlar todo lo anteriormente comentado, tecnologías como *Data Warehouse*, junto con la utilización de *Data Mining* y *Business Intelligence* en grandes organizaciones empresariales buscan obtener ventajas competitivas que marquen la diferencia frente a sus competidores en el mercado; esto nos presenta una dualidad entre el interés lucrativo que nos ofrece Big Data y, a la vez, la defensa de nuestra privacidad en el proceso productivo que hace posible las funciones anteriormente comentadas, de tal forma que este estudio tiene gran interés debido a los límites que debemos encontrar y establecer entre dichos dos puntos de interés, que aunque contrapuestos, están íntimamente ligados en la realidad, y se debe conocer que ninguna de las dos partes de la balanza que sostienen se incline demasiado.

Por todos estos motivos expresados existe un fundado entusiasmo personal en conocer cómo funcionan las herramientas que hacen posible Big Data y como se defienden los intereses relativos a la privacidad cuando se trabaja con los datos personales de los ciudadanos.

En cuanto a la metodología llevada a cabo para la revisión bibliométrica, fue llevado a cabo un proceso apoyado en estudios similares, revisiones bibliográficas, etc... que llevaron a cabo una metodología similar a la desarrollada en el apartado quinto de nuestro estudio (Domingo Carrillo, Chávez Miranda, & Escobar Pérez, 2016).

Realizar un estudio bibliográfico, incluyendo el máximo de artículos y demás material académico posible sobre la Privacidad en Big Data, de cara a obtener el mayor número de datos, cifras y documentos con los que investigar probabilidades, gráficas, tablas y conclusiones sobre el concepto principal (Congreso, Information, & Technology, 2016).

El sistema de búsqueda web en plataformas online, las plataformas web en las que encontrar los artículos y demás material académico utilizado, las herramientas utilizadas para el manejo de gran cantidad de datos y la forma de trasladar y exportar la información fueron aportadas tanto por los tutores académicos que supervisaron el mismo como por múltiples artículos académicos los cuales aportaron su metodología y dichos métodos se asimilaron lo mejor posible para el desarrollo de esta estudio (Congreso et al., 2016; Domingo Carrillo et al., 2016; Muñoz, Pedro, & Bolívar, 2013).

En cuanto al sistema de gráficas, tablas y demás material gráfico nos apoyamos en artículos que estudiaban el sector Big Data o su aplicación, y que realizaron un estudio bibliométrico mediante diferentes técnicas o formas, guiándonos en los espacios temporales utilizados y en los formatos de gráficas y tablas más utilizados (López-Borrull & Canals, 2013) por entender eran las referencias más correctas y productivas que se pueden plasmar en nuestro estudio.

Algunas ideas de desarrollo de contenido teórico fueron igualmente adaptadas de otros estudios, que si bien algunos de ellos no están íntimamente ligados con el objeto de estudio de nuestra investigación, fueron eficientes a la hora de estructurar el contenido teórico inicial del concepto (Froment, 2017).

Como podemos observar, el estudio bibliométrico de todos los contenidos académicos relacionados con un tema objeto de estudio, por comparación con otros estudios anteriores en determinados temas, entre los que se encuentra el Big Data y las tecnologías de la información y la comunicación, son el principal método para estudiar el estado de interés de un concepto y la cantidad de artículos y demás investigaciones productivas relacionadas con el mismo para desarrollar nuevas teorías. Por ello parece idóneo fundamentar la justificación de nuestro estudio en dichos métodos y estudios comentados anteriormente y tratar de plasmarlos de la manera más eficaz posible.

1.3. Objetivos

Con esta investigación se pretende esbozar un mapa de toda la investigación existente sobre Big Data y Privacidad archivada en las dos principales bases de datos a nivel mundial (Web of Science y Scopus), además de la base de datos bibliográfica de la Universidad de Sevilla (FAMA US).

De esta forma, con la identificación de la mayor parte posible de los artículos científicos que versan sobre Big Data y Privacidad, se pretende conocer la evolución que ha tenido la Privacidad de los datos personales en la aplicación de las herramientas más comunes utilizadas al aplicar Big Data a través de los años, los países y las universidades que crearon los artículos y libros encontrados y estudiados.

Se pretende adoptar una visión global del estado de cuestión del término, como objetivo prioritario, de tal forma que obtengamos un concepto global que nos permita vislumbrar si el mundo doctrinal, académico y popular tiene interés en un aspecto que tiene mucha cabida en nuestro día a día y si las proyecciones y el sentido común no nos fallan, en el futuro tendrá aún más prevalencia sobre nuestras rutinas, nuestras nuevas formas de ocio, nuestros nuevos ambientes de trabajo o incluso nuestras nuevas formas de entender la vida, y para ello es imprescindible conocer hasta qué punto podremos seguir siendo nosotros mismos, o lo que es lo mismo, tener nuestra intimidad en un mundo cada vez más globalizado y estandarizado.

Como objetivos secundarios plantearemos los siguientes:

- Realizar un estudio pormenorizado del estado de la cuestión actual de los términos Big Data y Privacidad, de cara a conocer cuánto interés real encontramos de los autores especializados en la materia de cara al desarrollo de artículos sobre el tema que ayuden a comprender mejor el mismo o a desarrollar nuevas habilidades y técnicas aplicables al mismo.
- Confeccionar un marco teórico conceptual que nos permita conocer y manipular los conceptos básicos que se utilizan en el tema en cuestión, existen muchos conceptos técnicos y herramientas especiales utilizadas en esta faceta, altamente vinculada a la informática, que se alejan de los conocimientos básicos de las ciencias económicas y expanden su área de influencia hacia otros ámbitos, tales como la ingeniería, la informática o la programación.
- Realización de un estudio bibliográfico y bibliométrico de artículos sobre el tema que atañe, mediante buscadores web de artículos académicos de conocido prestigio y su posteriores tratamiento, gestión y procesamiento de datos de todos ellos, plasmación de los mismos en tablas y gráficos que faciliten su estudio.
- Exposición de conclusiones alcanzadas gracias al conocimiento adquirido con el estudio global llevado a cabo, donde se procede a plasmar una conclusión por cada uno de los capítulos anteriormente mencionados y sirva de resumen para el conocimiento adquirido durante todo el proceso del estudio.

1.4. Metodología

Para la realización de este Trabajo Fin de Máster, basado en un estudio bibliométrico de la cantidad y calidad de los artículos basados en Big Data y la Privacidad de los datos de las personas en las herramientas que se utilizan en el mismo, para ello se ha llevado a cabo la selección y búsqueda de los artículos en las principales bases de datos (y sus respectivos buscadores) de Internet.

Una vez que se tienen en poder las listas de artículos, sus autores y todos los datos posibles, se lleva a cabo un volcado de datos en diferentes formatos sobre el programa gratuito Mendeley, muy útil en casos de gestión de referencias, artículos y libros académicos sobre Big Data.

Tras este paso se lleva a cabo una criba en la que, mediante la lectura de los resúmenes de dichos artículos, se eligen los artículos que más se amoldan al tema de interés buscado, dando como resultado un montante global final de artículos.

Para hacer más manejable y accesible el contenido de los mismos, se realiza una búsqueda web de cada uno de ellos, intentando acceder al documento digital original o copia, de cara a tener todo el contenido desarrollado (en su idioma original o traducido), así como la máxima información posible para su posterior estudio y desarrollo (en vista al análisis bibliométrico y a tener un grueso de información para poder desarrollar el marco teórico).

Tras esto, se procede a desarrollar una base de datos en el programa Microsoft Excel 2013, del paquete Microsoft Office 2013, en el cual se plasman los datos de los títulos, autores, año de publicación, país de publicación, universidad o institución productora, keywords, etc.... En general todos datos que se encuentran de todos los artículos posibles, de cara a tener una base de datos competente para sacar el máximo partido más tarde representándolo mediante gráficas, tablas, y demás contenido necesario.

En cuanto al contenido del marco teórico, toda la información se nutre de los materiales académicos descargados, formando un puzzle lo más homogéneo posible que permita obtener una visión del concepto y todas las influencias que puedan concurrir en el mismo.

De este modo se lleva a cabo la metodología de recogida de datos y utilización de los mismos, de cara a otorgar todo el contenido plasmado en el mismo la mayor credibilidad y eficacia posible en relación con la realidad.

Tras todo este proceso resultaron los siguientes apartados incluidos en el trabajo final:

1) Capítulo introductorio: integrado por introducción (breve resumen de intenciones y motivaciones para el estudio), justificación (donde se exponen las razones que motivan la posibilidad de llevar a cabo el trabajo y el interés del mismo), objetivos (fines básicos que alcanzar) y metodología (proceso de investigación llevado a cabo).

2) Marco Teórico: Los conceptos que se manejan en este estudio son bastante complejos, para una debida comprensión del mismo y poder manejar con eficacia y eficiencia toda la materia que a lo largo del mismo se desarrolla es necesario dotar al lector de unos conocimientos mínimos que otorguen una visión capacitada.

3) Estado de la cuestión: La situación actual de los términos más utilizados y estudiados en el trabajo son de vital importancia para adquirir una visión global de la popularidad del mismo, la situación en la que se encuentra en la actualidad y la cantidad de estudiosos que se interesan por el desarrollo del ámbito aplicativo.

4) Otros estudios similares: Para dotar de fundamento y profundidad a lo explicado en el anterior capítulo, se desarrollan a modo de ejemplo el contenido e ideas que se reflejan en estudios parejos al que se desarrolla en este texto, de cara a conocer de primera mano ideas, conceptos, formas de aplicar y desarrollar metodologías similares al campo de estudio, se pretende dotar al lector de una visión más aproximada de la cantidad y calidad de material que se desarrolla alrededor del termino objeto de estudio.

5) Análisis de datos y resultados: Conjunto de datos, gráficas y tablas de información resultados del proceso de investigación llevado a cabo, con el desarrollo argumental de cada uno de ellos, las explicaciones debidas, razonamientos lógicos aplicados y clasificaciones de autores, revistas y países productores sobre la materia.

6) Conclusiones: Decisiones o consecuencias fruto del estudio, desarrollo y experimentación llevados a cabo mediante análisis bibliométrico del término y su repercusión actual, dando como resultado una serie de conceptos definitivos que se pueden obtener al revisar y analizar los datos obtenidos, así como la redacción de los límites y problemas encontrados y las posibles vías de investigación futuras.

La metodología, tratamiento y alcance de los datos obtenidos se analiza y desarrolla con una mayor precisión en el capítulo quinto de este estudio.

CAPÍTULO 2. Big Data y Privacidad.

2.1. Big Data

En este apartado procedemos a realizar un análisis del concepto Big Data, así como de sus principales tipos, características, ciclo de vida y herramientas con las que se lleva a cabo.

2.1.1. Concepto

Hoy en día se asiste a una digitalización de todo lo relacionado con las actividades cotidianas del ciudadano, en todas las facetas de la vida, desde el ámbito social (redes sociales), pasando por las vías de comunicación en la red (portales webs de noticias) e incluso en el ámbito de ocio y diversión (videojuegos, tanto en plataformas móviles como sobremesa).

Todo esto ha conllevado la creación y el flujo activo de un enorme volumen de datos, tanto personales como en masa, a lo largo y ancho de Internet. Dicho flujo de datos es lo que se ha denominado “Big Data”, y no dejaría de ser un mero hecho sintomático de las nuevas formas de relación y ocio de los tiempos que nos ocupan, si no fuese porque las empresas y organizaciones gubernamentales gozan de las herramientas necesarias para hacerse con dichos datos, procesarlos, sacar información valiosa de ellos, sacar conclusiones gracias a la información que les aporta y tomar decisiones en pos de las mismas (Kitchin, 2015).

Ejemplos de dichas decisiones pueden ser estrategias de marketing enfocadas a un cliente o sector del mercado en concreto, tratar de comprender el comportamiento o el enfoque que toman una serie de clientes o incluso detectar amenazas presentes y futuras en el desarrollo de un comercio, las posibilidades son tan grandes como podamos imaginar y tan útiles como seamos capaces de conformar (Mittelstadt & Floridi, 2016).

El procesamiento y utilización de dicha información se hace posible para las empresas a través de lo que se llama “Big Data Analytics”, el conjunto de herramientas y técnicas que hacen posible el trato con estos grandes volúmenes de datos, puesto que con las técnicas de bases de datos convencionales no sería posible (Jensen, 2013).

Todo esto presenta la parte positiva del asunto, encontrándose ante un concepto en desarrollo e investigación que permite un rango de mejora excelente en el futuro, pero por el camino encontramos una pequeña deficiencia con la que se debería reparar y tener mucho cuidado.

Resulta que, mientras esta herramienta supone una inmensa oportunidad para empresas y organizaciones, en la otra cara de la moneda se encuentra un riesgo alto ante la privacidad e intimidad de todas las personas, entidades y empresas que se encuentran integradas en los datos que se manejan en masa, de tal modo que quede la información personal al descubierto de los intereses privados de las empresas. Las herramientas de análisis que se utilizan para llevar a buen término el Big Data consisten en almacenar, gestionar y analizar eficientemente diversos datos recogidos de todas las fuentes posibles y disponibles (van de Pas & van Bussel, 2014).

La consecuencia de esto deriva en que las personas se encuentran en una circunstancia de máxima vulneración de su intimidad, debido a que se encuentran completamente vulnerables a la exposición y exploración ajenas de datos personales y de comportamiento; gracias al Big Data es posible obtener más datos que con herramientas convencionales y ello puede producir violaciones de seguridad y privacidad si dicha información recae en individuos o instituciones que no manejen la misma de manera adecuada (Sommer, 2015).

Y ante esta situación, el estudio de este Trabajo Fin de Máster trata de estudiar los conceptos y características fundamentales del concepto Big Data, y una vez establecido el estudio del concepto, tratar de estudiar las técnicas de protección de los datos que se manejan más comúnmente en las herramientas que hacen posible el Big Data, llevando a cabo para ello un estudio bibliométrico de los principales artículos sobre el tema, de los cuales se obtiene dicha información.

2.1.2. Tipos de información manejados en Big Data

De cara a ofrecer una forma reglada de manejar las grandes cantidades de datos en Big Data, es necesario desde un primer momento comprender que no todos los datos se almacenan ni operan igual, debido a sus diferentes características, formas de acceder a ellos, formas de procesarlos, etc... Por ello es común que en muchos estudios relativos al tema que nos ocupa se encuentren clasificaciones de diferentes formas de los datos que se manejan en las herramientas de procesamiento de datos en Big Data, dichas clasificaciones varían enormemente de un ámbito de aplicación a otro (son muy diferentes los datos manejados en el área sanitaria que en área militar o publicitario, por ejemplo), pero en líneas generales se distinguen los siguientes tipos:

- A) Datos estructurados: Se trata de todo tipo de datos que se puedan modelar, introducir, almacenar, consultar, procesar y visualizar de forma cómoda y asequible. Suele estar presentado como campos predefinidos, con tipos y tamaños específicos, gestionados mediante bases de datos relacionales o en hojas de cálculo (B, Wu, & Jiang, 2016). Estos datos facilitan el análisis y la extracción de información útil, ya que el procesamiento no requiere equipos de muy alto rendimiento o técnicas de análisis complejas (Pérez, 2015).
- B) Datos semiestructurados: Se pueden considerar como un subgrupo de los datos estructurados, pero no siguen un modelo rígido (Jain, Gyanchandani, & Khare, 2016). Además de seguir el esquema que amolda toda la estructura de los datos, suele ir acompañado de etiquetas o marcadores que se utilizan para identificar elementos excepcionales de cada elemento que los hacen diferenciables al resto, suelen ostentar una representación jerárquica de diferentes campos dentro de la estructura de datos que presentan (Mishra & Singh, 2016).
- C) Datos no estructurados: Se trata de registros presentados y almacenados sin un formato o esquema predefinido. Podemos encontrarnos en este apartado con texto de formato libre como pueden ser libros o artículos, artículos científicos como la bibliografía que compone este Trabajo Fin de Máster o incluso archivos de imagen, audio o video (J. Zhang, Yang, & Appelbaum, 2015). La dificultad que presentan estos datos se encuentra en que no es posible representarlos de una manera rígida como los dos anteriores, lo que ha llevado a introducir nuevos mecanismos de procesamiento para trabajar y manejar este tipo de datos, los cuales analizaremos más adelante (Sudarsan, Jetley, & Ramaswamy, 2015).

2.1.3. Características de los datos de Big Data

Se debe tener en cuenta que el concepto de Big Data a día de hoy se encuentra en una fase que podríamos considerar aun embrionaria, por lo cual las características del mismo se encuentran en continua evolución y modificación, pero actualmente, y habiéndose analizado una gran cantidad de artículos de investigación en relación con el concepto Big Data, la mayoría de los autores coinciden en los siguientes:

- A) Volumen: La primera característica, y por la cual recibe su nombre, es la cantidad de datos que se encuentran para ser analizados en Big Data, normalmente dicha cantidad se encuentra escalada al término "Exabyte". Cuanto más grande sea dicho volumen, más significativo será para el procesamiento y su análisis (Tascón, 2013).

B) Variedad: Uno de los criterios más importantes de los datos tratados en Big Data es la variedad del contenido (Jain et al., 2016). Al igual que el contenido puede ser estructurado o no estructurado, también se puede dividir en interno o externo. Los datos internos podemos definirlos como los datos que maneja la empresa u organización dentro de su entidad, es decir, mediante sistemas ERP o CRM. Los datos externos se recopilan mediante fuentes de información ajenas a la actividad organizativa de la empresa u organización, ejemplos de ello pueden ser la propia Internet, bases de datos relacionales de pago, etc... Además de ello en los tipos de datos encontrados se pueden incluir atributos textuales, numéricos y binarios (Gheid & Challal, 2015).

La variedad es importante debido a que, cuanto mayor variedad tenga la base de datos de la que se extrae la información, estadísticamente mayor será la variedad de datos que se logren extraer. Esta característica se encuentra altamente ligada con la siguiente.

C) Veracidad: Es la característica de mayor importancia, los datos recolectados y tratados deben tener una exactitud y validez máximas, de no ser así, más tarde los datos arrojados no serán un reflejo fiel de la realidad, y se estarán obteniendo falsas conclusiones sobre las que adoptar nuestras decisiones. Por ello es de vital importancia que a la hora de recoger dicha información se haya seguido un proceso estricto de organización de registros y se haya puesto mucho énfasis en eliminar cualquier duda que pudiesen arrojar los datos recogidos y que no garantizaran su integridad y valor (Yoon, Hoogduin, & Zhang, 2015).

D) Valor: Las tres características anteriores son muy importantes, evidentemente, pero no debemos olvidar el fondo del asunto que se trate de investigar mediante técnicas Big Data, y ello no es nada más y nada menos que el valor que dichos datos tengan en su interior. Es decir, si se procede a analizar el mercado de la venta de coches para sesgar publicidad de la forma más efectiva posible, de poco valdrá tomar una base de datos impoluta sobre, por ejemplo, el número de turistas que visitaron las playas de Andalucía durante el año 2016, quizás dicha base de datos cumpla con los requisitos de gran volumen, veracidad y variedad, pero su valor no aporta nada al estudio que se trata de llevar a buen puerto (Bolognini & Bistolfi, 2016).

E) Velocidad: Entendiendo este último criterio como la velocidad a la que se generan nuevos datos que interfieran en los datos ya recogidos y la manera en la que cambian los mismos cuando ello ocurre (Gheid & Challal, 2015). Big Data no sólo se basa en datos estáticos, más bien todo lo contrario, los datos a los que somete están continuamente influidos por cambios en tiempo real, por lo cual obliga a la extracción de resultados rápidamente cuando tratamos de adoptar decisiones drásticas (Kshetri, 2014).

2.1.4. **Ciclo de vida de los Datos**

- 1) Adquisición y pre procesamiento de datos: Debido a la diversidad de tipos de medios desde los que se importa información (texto, imagen video, página web) y la gran cantidad de tipos de datos que se pueden encontrar (estructurados, no estructurados, etc...), el primer paso del ciclo de vida de los datos a llevar a cabo es adquirirlos y pre procesarlos, de tal manera que queden unificados y gocen de alta calidad para obtener unos resultados competentes y fiables en el procesamiento posterior. Es indispensable eliminar datos similares, repetitivos o que no aportan ningún tipo de interés o valor al estudio o trabajo que estemos llevando a cabo (X. Li & Shen, 2016).
- 2) Análisis de los datos: Se trata del proceso central de la aplicación de Big Data. Se puede dividir en 3 categorías: arquitectura informática, consulta e índice, análisis y procesamiento de datos (X. Li & Shen, 2016). Otros estudiosos del término dividen este apartado en dos, "Model planning" y "Model execution" (Robak, Franczyk, & Robak, 2016).

En la planificación del modelo el equipo debe decidir qué métodos, técnicas y flujos de trabajo se van a utilizar, se deben seleccionar las variables clave que más afecten o hagan más productivo al modelo, establecer dependencias entre las variables, hasta finalmente optar por los modelos de procesamiento de datos más apropiados y más productivos para el trabajo que se tiene entre manos y la finalidad a la que se quiere llegar. En la fase de construcción del modelo, se deben desarrollar los modelos previamente fijados en la fase anterior, es usual que en esta fase se lleven a cabo modificaciones y adaptaciones necesarias de hardware y software, de cara a obtener unas herramientas más rápidas, fiables y productivas (Robak et al., 2016).

- 3) Comunicación de resultados y operatividad de los mismos: En esta fase se expondrían/comunicarían los resultados para su debida explotación por parte de la empresa, se debe otorgar un resumen de los resultados obtenidos en el proyecto, como pueden ser decisiones importantes a adoptar a modo de conclusiones o algún tipo de cuantificación del negocio llevada a cabo. En la última fase de “operacionalización” de los datos, se deben entregar los resultados finales del proyecto, ya sea mediante informes finales, documentos o presentaciones. El valor comercial y de conocimiento del proyecto debe ser adecuadamente transmitido a las partes interesadas del mismo.

Cabe decir que todas estas fases mencionadas anteriormente pueden mezclarse entre sí, e incluso llevarse a cabo simultáneamente, pueden avanzar hacia delante y hacia atrás, son tremendamente volubles y flexibles, ya que se adaptan en tiempo real a las necesidades que se encuentren, los problemas acaecidos a la hora de recopilar la información o tratarla y la profesionalidad y capacidad del equipo de trabajo con el que se cuente (Robak et al., 2016).

2.1.5. Herramientas de Big Data Analytics

Big Data, como se ha mencionado, bien implantado y llevado a cabo, puede conllevar una verdadera revolución dentro de una empresa u organización, pero para ello se necesitan una serie de herramientas que sean capaces de trabajar con datos que cumplan dichas características:

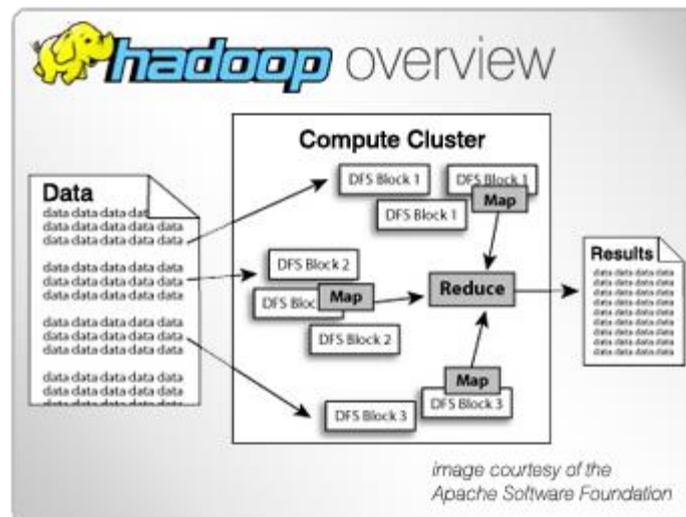


Ilustración 1: Forma de trabajar de Hadoop

Fuente: Dataprix (2016), recuperado de: <http://www.dataprix.com/software-empresa/apache/apache-hadoop>

Apache Hadoop: Es la herramienta más popularizada y utilizada, se trata de una herramienta analítica de código abierto concebido para manejar datos de Big Data. Cuenta con una arquitectura de datos no relacional que puede manejar grandes cantidades de datos heterogéneos distribuidos. Mediante el almacenamiento de datos en nodos distribuidos, procesa los datos y realiza cálculos sobre ellos con la siguiente herramienta, que podríamos considerar incluida dentro de Hadoop (Tascón, 2013).

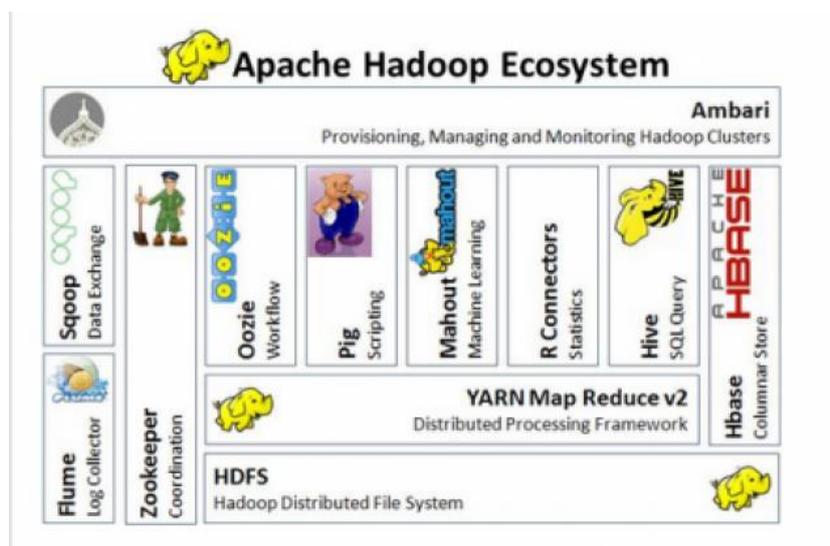


Ilustración 2: Ecosistema Hadoop

Fuente: Dataprix (2016), recuperado de: <http://www.dataprix.com/software-empresa/apache/apache-hadoop>

A continuación se procede a definir y estudiar las principales herramientas integrantes de Apache Hadoop, que si bien funcionan independientemente y cumplen sus funciones eficazmente, trabajan codo con codo para hacer funcionar al sistema y conseguir los resultados y análisis de datos perseguidos.

HDFS: Son las siglas de "*Hadoop Distributed File System*", se trata del núcleo procesador de Hadoop. Es un sistema de archivos distribuido para el almacenamiento y la gestión de almacenes de datos masivos. Su forma de trabajar se podría resumir en la creación de divisiones de datos mediante bloques, y la asignación de dichos bloques a diferentes servidores dentro de la propia aplicación, para permitir su procesamiento posterior, es un archivador de datos en forma masiva, por resumirlo de alguna forma.

PIG: PIG es una plataforma que permite desarrollar programas *MapReduce*. El lenguaje de programación se llama '*PIG Latin*'. Su objetivo es aumentar el rendimiento de Hadoop y MapReduce ofreciendo un lenguaje de programación que permite un procesamiento más rápido (Mishra & Singh, 2016).

Mahout: Se trata de un software desarrollado sobre Hadoop, dentro de su interior se encuentran un conjunto de algoritmos que se encargan de filtrar, clasificar y agrupar los conjuntos de datos distribuidos. Según algunos autores otorga una biblioteca de minería de datos y aprendizaje escalable, para lograr una privacidad más alta que utilizando MapReduce simplemente, y además de ello preservando la minería de datos o análisis (X. Zhang et al., 2015).

Hive: Se trata de una herramienta de data warehouse que permite solicitar y administrar grandes cantidades de datos distribuidos. Hive proporciona la capacidad de acceder al almacenamiento utilizando un lenguaje informático similar al SQL denominado HiveQL.

MapReduce: Se trata de una herramienta creada por Google para procesar grandes cantidades de datos. Como bien nos plantea su propio nombre, su forma de trabajo consiste en “mapear”, que se traduce en distribuir los datos en varios clústeres para el procesamiento en paralelo, y además de ello “filtra” todos los sub-resultados para proporcionar una retroalimentación final, la distribución no afecta a la calidad de los datos anónimos (Zakerzadeh, Aggarwal, & Barker, 2015).

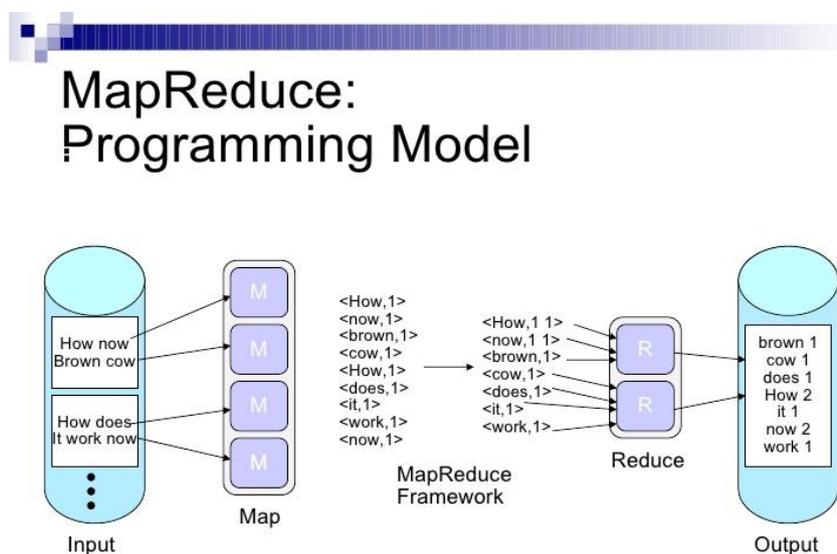


Ilustración 3: ¿Cómo trabaja MapReduce?

Fuente: Sri Prasanna (2009), recuperado de: <https://www.slideshare.net/sriprasanna/map-reduce-1941647>

Actualmente, gracias al procesamiento de datos a gran escala, se han desarrollado aplicaciones para la aplicación de Big Data. Se obtiene una herramienta más potente con MapReduce si lo combinamos con los recursos de una infraestructura “en la nube”. Empresas de renombre han sacado al mercado paquetes de esta herramienta y los han puesto a disposición de cualquiera que pueda pagarlos para aplicarlos, como puede ser Amazon y su *Amazon Elastic Mapreduce service* (o Amazon EMR, sus siglas comerciales) (X. Zhang et al., 2015).

NOSQL: Bases de datos que no son relacionales. Proporciona la capacidad de consultar y recuperar datos no estructurados o semi-estructurados (Mishra & Singh, 2016). La infraestructura de seguridad que aporta esta herramienta sigue evolucionando día a día debido a lo altamente popularizado que se encuentra; a pesar de ello, los desarrolladores informáticos que utilizan esta herramienta normalmente incorporan seguridad mediante otros medios a sus proyectos de trabajo, esto es debido a que en la creación de estas bases de datos se tuvieron en cuenta diferentes problemas e intereses que no abordaban inicialmente la temática de la seguridad de los datos como principal objetivo, y la protección de los mismo ha debido integrarse más tarde, mediante modificaciones o con programas externos que se encargan de dicha función. En general, las compañías que adoptan NoSQL en su actividad confían la seguridad de sus proyectos a mecanismos de aplicación externos (CSA Cloud Security Alliance, 2012).

A continuación se procede a definir una de las herramientas de Bases de datos NoSQL mas utilizadas:

Cassandra: Se trata de una base de datos NoSQL orientada por columnas, se ha vuelto muy conocida gracias a su capacidad para facilitar el acceso a datos con un gran volumen de registros, debemos destacar que es una herramienta analítica que ha sido desarrollada por Facebook, y se encuentra altamente popularizada entre los trabajadores de Big Data. Su nombre está inspirado en un personaje de la mitología griega que gozaba del don de la profecía. Entre sus características encontramos que es distribuida, escalada linealmente (lo cual la dota de predictibilidad), escala de forma horizontal (se pueden añadir nuevos nodos facilmente) e implementa una arquitectura peer-to-peer (una forma muy eficiente de gestión de datos internos del sistema). A pesar de todas estas ventajas, algunos autores defienden que no es una herramienta adecuada para alojar un data warehouse convencional (Zaforas, 2016).

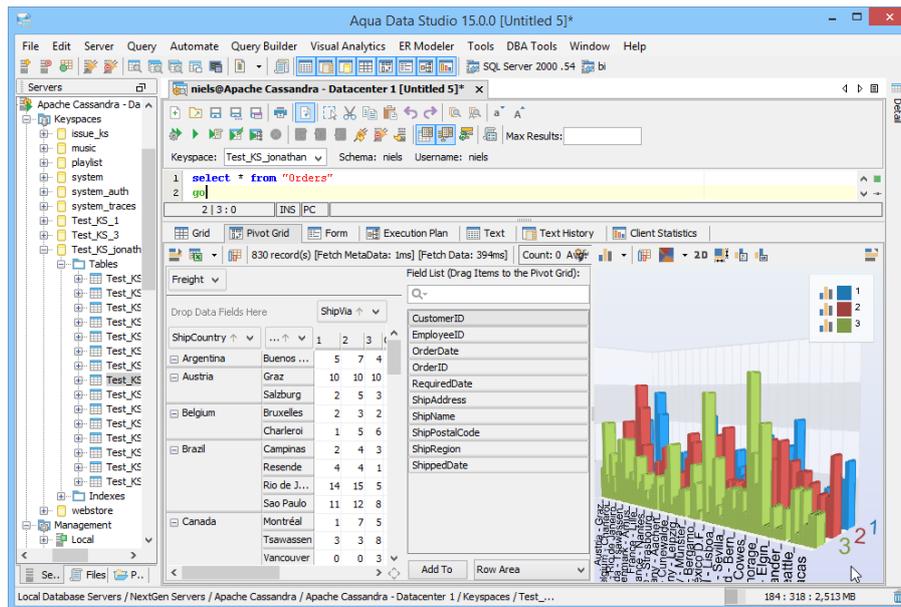


Ilustración 4: Interfaz de Cassandra aplicándose gracias al soporte de Aqua Studio

Fuente: Aquafold (2013), recuperado de:

<https://www.aquaclusters.com/app/home/project/public/aquadatastudio/wikibook/New-Features-14.0/page/Apache-Cassandra-Support/Apache-Cassandra-Support>

2.2. La privacidad de la información en el Big Data

Se procede en este capítulo a estudiar el concepto y características de la Privacidad dentro del ámbito aplicativo de Big Data.

2.2.1. Concepto de Privacidad

Se podría definir la Privacidad como: reclamación de individuos, grupos o instituciones para determinar por sí mismos cuándo, cómo y en qué medida la información sobre ellos se comunica a otras personas ajenas; la Privacidad es el interés que tienen los individuos en mantener un "espacio personal" (Perera, Ranjan, & Wang, 2015).

Así mismo, se pueden encontrar otras definiciones y reconocimientos de la Privacidad en diferentes textos normativos a escala internacional, como por ejemplo en la Declaración Universal de Derechos Humanos, la Convención Europea, reflejado en el artículo 7 de la Carta de los Derechos Fundamentales de la Unión Europea donde se define de la siguiente forma "el respeto a la vida privada y familiar", incidiendo en el tema y otorgándole aún mayor protección en el siguiente artículo 8, específicamente relativo a "la protección de derechos personales".

También se encuentra reflejado el concepto en el artículo 12 de la Declaración Universal de Derechos Humanos, protegiendo a todos los individuos de cualquier “interferencia arbitraria contra su privacidad, familia, hogar o correspondencia, así como ataques al honor y a la reputación.

Para que se respete la Privacidad de los datos en Big Data, algunos autores (Mantelero, 2015) defienden que deberían cumplirse los siguientes requisitos para darse adecuadamente defendida: los propietarios de los datos deben saber a qué tipo de datos se accederá, qué tipo de fuentes de datos se combinarán, dónde se procesarán los datos, qué tipo de análisis se utilizará, qué tipo de resultados se generan, y así sucesivamente. Incluso más, los propietarios de datos deben ser capaces de intervenir en cualquier momento durante el ciclo de vida de datos para que puedan retirar o cambiar su consentimiento con el tiempo. Si se cumplen todos estos requisitos, los propietarios de datos deben tener control sobre sus datos en tiempo real (Perera et al., 2015).

2.2.2. El desafío de la seguridad y la Privacidad.

Big Data es una tecnología novedosa, debido a su modo de aplicación y la eficacia de los datos y resultados que otorga, las empresas y organizaciones públicas que conocieron de su existencia las implantaron en sus actividades ordinarias rápidamente, de cara a predecir las tendencias del mercado o el comportamiento de los usuarios, por poner un ejemplo, otorgándoles ventajas importantes de cara a la competencia o a la eficacia de sus actividades ordinarias.

Las herramientas de analítica de datos anteriormente mencionadas fueron creadas por matemáticos e informáticos que pusieron mucho hincapié en su eficacia de extracción y resolución de datos en volúmenes ingentes, pero no prestaron demasiada atención a otro tipo de características que influían en el cometido de todas esas herramientas, la seguridad y la protección de la privacidad de los datos personales.

Las empresas, una vez aplican Big Data en sus análisis, optan por fijarse en los beneficios que les otorgan y en las ventajas competitivas que adquieren al tomar decisiones correctas con este instrumento, pero pueden no atender a la privacidad de los datos masivos que manejan. Y ello nos plantea las siguientes problemáticas:

- A) **Distribución aleatoria:** El concepto sobre el que pivota el Big Data se basa en el paralelismo, para lo cual los datos masivos se almacenan y procesan en servidores distribuidos y que actúan como una herramienta compacta bastante eficiente. El problema de este “modus operandi” reside en que es muy difícil localizar el lugar exacto de almacenamiento y procesamiento de cada uno de los datos que se procesan, y ello puede dar lugar a problemas de seguridad legal. El principal problema en el aspecto de la Privacidad es, pues, poder establecer una distribución de los datos para su procesamiento de una forma que se adecuen a las regulaciones vigentes y sean propicias con la sensibilidad de los datos que manejan.
- B) **Privacidad:** Otro gran desafío a suplir por las tecnologías aplicables en Big Data es el que trata a los datos sensibles con la atención e importancia que se merecen. Actualmente se tratan todos los datos con la misma prioridad (cuando evidentemente no se le otorga el mismo valor al número de nuestro DNI o dónde se encuentra su residencia habitual a el número de veces que ha visitado el cine en el último mes), sin asociar a ciertos datos sensibles un cifrado concreto o el llamado “procesamiento ciego”; de tal forma que cualquier organismo o persona propia con una actitud maliciosa puede alterar dicha información en los datos o robar información de la misma, realizando un daño irreparable a la persona privada de la que nace o se sustrae dicha información.
- C) **Cálculo:** La idea que aboga el Big Data se centra en la extracción de información mediante la aplicación de cálculos específicos y eficaces. Pero igual que es importante llevar a cabo dichos cálculos, también lo es blindar y proteger la forma de realizarlo para evitar los riesgos de sesgar o modificar los resultados finales al libre albedrío de un agente externo interesado en ello. De tal modo, es importante proteger las herramientas de extracción de datos de Big Data de dichos agentes, para evitar los intentos de espiar la naturaleza de los datos.
- D) **Integridad:** En Big Data, que exista una gran cantidad de datos no tiene que conllevar siempre que los mismos den unos resultados de calidad. Por ello, antes de comenzar a trabajar con bases de datos masivas, es fundamental que se garantice la validez y el nivel de confianza depositados en dichos datos, se debe evitar confiar y trabajar sobre registros sospechosos o dudosos.

- E) **Comunicación:** La forma en la que se organizan los datos en Big Data por los programas informáticos habilitados para ello es bastante compleja y concreta. Si una persona exterior a la empresa o el organismo que trata con los datos descubre una forma de modificar la comunicación entre los nodos de trabajo (anteriormente mencionados, clústeres en servidores, etc...) podría poner a información muy valiosa en manos de personas ajenas. Por lo tanto, es un buen reto para las grandes herramientas de datos adoptar nuevos protocolos de red seguros para proteger las interacciones entre las diferentes partes.
- F) **Acceso a la información:** Por último, el acceso a la información en Big Data debe estar fuertemente respaldado por métodos y herramientas que garanticen y denieguen el acceso a personas ajenas que no tengan la responsabilidad de tratar con los datos. Cualquier modificación en el estado de los clústeres o servidores como la creación o la supresión de nodos debe ser supervisada por un mecanismo de autenticación para proteger el sistema (Mishra & Singh, 2016).

2.2.3. La seguridad en los métodos analíticos de Big Data

Debido a la eficacia y la utilidad de Big Data en los procesos de las empresas y las decisiones empresariales, es lógico pensar que muchas de ellas hayan decidido adaptarlas a sus necesidades en cuanto conocieron de su existencia, el problema es que al centrarse en la utilidad y eficacia de los mismos, se dejaron de lado los aspectos relacionados a la seguridad y la privacidad de los datos que se manejaban, o al menos no se le prestó toda la importancia que se debiera. En este apartado trataremos de estudiar las contribuciones más interesantes en busca de una mejor privacidad y seguridad en los datos que se manejan en Big Data.

Algunos autores (M. Li, Zang, Bai, Yu, & Liu, 2013) presentan soluciones ante este problema; este en concreto nos propone una arquitectura que permite proteger los problemas de Privacidad en entornos de almacenamiento en la nube y Big Data. Esta herramienta se encargar de inspeccionar y rastrear la actividad de los usuarios, reduciendo la autoridad de los nodos menos fiables; además es personalizable, permite a los propios usuarios del almacenamiento en nube configurar su protección de la Privacidad, también permite reducir la capacidad del proveedor de modificar la configuración de privacidad.

Tanto la transparencia como la privacidad son elementos esenciales en una sociedad democrática actual. En el ámbito de las administraciones públicas, la transparencia debe asegurar que las acciones y decisiones adoptadas por el gobierno pueden ser analizadas al completo, mientras que la privacidad asegura que los ciudadanos no pueden ser monitoreados y son libres de expresar cualquier opinión, por diferente que sea.

La transparencia debe permitir evitar y delimitar el abuso de poder por parte del gobierno y asegurar que los ciudadanos y las empresas sean tratados por igual. Es esencial que las decisiones del gobierno sean transparentes y los ciudadanos deben ser conscientes de lo que se está decidiendo y cuánto dinero se gasta, con qué propósitos. Esto, a su vez, debería dar lugar a una confianza creciente en los que están en el poder (Janssen & van den Hoven, 2015).

La Privacidad es un fundamento de nuestra democracia, porque sin la Privacidad las personas pueden ejercer poder para influir en los demás. Esta es una de las razones por las que el gobierno está organizado en dominios y la información se almacena en diferentes departamentos independientes, de cara a evitar la situación que, por ejemplo, un funcionario público tenga acceso a todos los datos de un ciudadano, sin tener restricciones o limitaciones de cómo tratar o hasta dónde poder llegar con dicha información. Esto a menudo se basa en la idea de que la información reunida para un propósito podría ser usada para otros propósitos diferentes y ajenos a los iniciales (Janssen & van den Hoven, 2015).

2.2.4. Posibles Técnicas para proteger la Privacidad

Una plataforma de Big Data se compone de varias evoluciones tecnológicas, tanto si hablamos de capacidad de almacenamiento como potencia de procesamiento de los datos que contiene. Por lo cual, las técnicas de seguridad tradicionales no pueden ser eficientes aplicadas a Big Data, pues estas trabajan a otro nivel. Se hace necesario, pues, estudiar y crear nuevas técnicas de seguridad que acompañen la evolución de la tecnología y las necesidades empresariales.

Las soluciones deberán ser equilibradas entre regulaciones, controles de seguridad y análisis de datos. En los siguientes apartados, proponemos un conjunto de técnicas que ayudan y podrían ser utilizadas como base para asegurar la Privacidad en Big Data.

2.2.4.1 Normas y legalidad

Los Big Data son un fenómeno enorme que continúa cambiando el mundo ya que se convierte en una fuente importante y potencial para tomar decisiones en las empresas. Las organizaciones y los gobiernos se están beneficiando de la recolección, análisis y procesamiento de grandes volúmenes de datos de todas las fuentes posibles para extraer información valiosa.

En el ámbito internacional europeo nos encontramos con la Directiva sobre la Privacidad de la Unión Europea 95/46 / CE (1995), que protege a las personas en lo que respecta al tratamiento y la transmisión de datos personales, existe desde los últimos años del siglo XX. Fue modificada por la Directiva 97/66 / LC (UE 1997), ampliando el ámbito de aplicación a los servicios electrónicos y finalmente sustituida por la Directiva sobre la Privacidad y las comunicaciones electrónicas (UE, 2002) (van de Pas & van Bussel, 2014); como podemos observar, la legislación de aplicación en el ámbito europeo se encuentra desfasada temporalmente, encontrando que la legislación más avanzada fue formalizada y publicada hace 15 años, y por motivos evidentes los medios tecnológicos son diferentes en la actualidad a los que encontrábamos entonces.

Hoy en día, Google, Apple, Facebook, Twitter y Amazon, son los mayores agregadores de datos en todo el mundo y obtienen de sus diferentes herramientas y productos gran cantidad de datos confidenciales de los ciudadanos que utilizan sus servicios. Por ello, estas empresas deben dejar lo más claro y transparente posible para el usuario cómo sus datos están siendo manejados, almacenados y a las personas y medios que les serán revelados y tendrán acceso a ellos.

La asimetría de poder entre los usuarios y los proveedores de servicios genera riesgos para la privacidad de los usuarios, ya que los proveedores de servicios están en mejores condiciones de utilizar la información que consiguen de la forma que más atienda a sus intereses (van de Pas & van Bussel, 2014), la idea que se pretende establecer es la de que exista un dialogo continuo entre los proveedores de servicios y los usuarios de los mismos, creando un ecosistema “vivo”.

En el cual existan continuamente las mejores garantías de respeto hacia la privacidad del usuario, que las necesidades del servicio se adapten de la mejor forma a los intereses de ambas partes y que, por supuesto, no se viole la intimidad de ningún usuario ya que todos ellos conocen claramente hasta qué punto son tratados los datos que incluyen en el programa del proveedor de contenidos.

Sin embargo, a nivel europeo y sobre todo a nivel nacional, se encuentra la situación de que no existen suficientes leyes y normas legislativas que regulen la minería de grandes volúmenes de datos y la privacidad de los usuarios en ella. Como esta última recopila y almacena todos los registros digitales, puede contener información financiera, de salud o información sensible de cualquier ámbito. Es decir, no todos los datos almacenados podrán ser procesados y extraídos para obtener información. Además, como los datos masivos que se manejan en Big Data funcionan bajo un modo de almacenamiento distribuido en todo el mundo, es importante elegir cuidadosamente el lugar de almacenamiento en el que depositaremos toda esa ingente cantidad de información, de cara a su protección.

Por lo tanto, se debe desarrollar un conjunto de leyes y reglamentos que doten al Big Data de un bagaje necesario para hacerlo un beneficio y una tecnología segura de utilizar y de aplicar al mundo real, teniendo en cuenta todas las características anteriormente mencionadas, y otorgando la mayor protección y seguridad posible a los datos manejados.

Se ha de citar aquí la necesidad de que las empresas, pero sobre todo la Administración Pública y las universidades abran sus bases de datos y permitan su manejo de forma abierta a los ciudadanos o compañías que quieran hacer uso de esos paquetes de información, siempre y cuando no sean datos sensibles y estén sujetos a una serie de limitaciones y normas. Si los datos son el petróleo del siglo XXI, ese combustible no sirve de nada si no podemos extraerlo. Es obligación de nuevo de los legisladores establecer el marco en el que esto ha de darse y respetarse. España es el único país de la Unión Europea que no cuenta con una Ley de transparencia, que no ha desarrollado una legislación que obligue a poner a disposición ciudadana aquellos datos que, no afectando a la privacidad y seguridad, los ciudadanos tienen derecho a poder conocer si los necesitan (Tascón, 2013).

2.2.4.2 Encriptación o cifrado

El cifrado es siempre una buena técnica para proteger los datos sensibles. En el caso de datos masivos, el cifrado podría ser perfectamente empleado para asegurar diferentes componentes tales como:

- Almacenamiento: En Big Data generalmente se almacenan los datos en los servidores dedicados a ello tal y como están y sin un procesamiento previo. Además, los datos sensibles y no sensibles se tratan de igual forma. Por lo tanto, si un tercero con intenciones dañinas accede al almacenamiento, los datos privados pueden ser fácilmente encontrados y sustraídos.

Por lo tanto, todos los datos (o al menos los más sensibles) deben almacenarse en forma encriptada, de tal forma que si alguien externo consigue tener acceso al almacenamiento, no pueda extraer ningún conocimiento de la base de datos, ya que sólo los usuarios que poseen las claves de descifrado pueden acceder a la misma. Para evitar el acceso a este almacenamiento de datos a gran escala, algunos autores han desarrollado soluciones a modo de algoritmos de computación [ORAM algorithm (P. Li & Guo, 2014)], que tratan de conseguir mayor privacidad y seguridad.

- Cálculos: Es muy difícil predecir o averiguar donde se realiza el procesamiento de datos en Big Data, por lo cual se hace tremendamente necesario impedir que terceros ajenos a la empresa o la organización puedan acceder a los resultados de procesamiento. Para ello se han creado las técnicas de procesamiento ciego basadas en cifrado homomórfico (Perera et al., 2015), que permiten a un procesador ejecutar un conjunto de instrucciones sin quedar grabada la naturaleza de los datos.
- Comunicación: La comunicación entre las partes que se puedan dar en Big Data debe estar protegida en todo momento para evitar posibles intrusos extraños. Una solución eficaz es encriptar las redes de comunicación empleando protocolos de seguridad (las más populares son SSL o TLS), mediante las cuales las comunicaciones son ilegibles para terceros ajenos a la empresa que no conocen las claves de entrada a los datos.

2.2.4.3 Autenticación:

Los mecanismos de autenticación son una forma eficaz de controlar el acceso a los datos de todo tipo. Las tecnologías que manejemos de Big Data deben emplear alguna técnica de este tipo para controlar el acceso, unir los servidores de los que se dota la información y controlar la información de los datos más sensibles (Mahalakshmi, n.d.).

2.2.4.4 Metadatos:

Otra posible solución para diferenciar los datos recogidos según su importancia y la posibilidad de incluir en el tratamiento o no es la implicación de metadatos y técnicas de etiquetación de la información. De esta manera, las grandes herramientas de datos no tratarán todos los registros de la misma forma y sería posible respetar la información privada durante el procesamiento de los datos masivos (Mishra & Singh, 2016).

2.2.4.5 Distribución desestructurada:

Debemos evitar almacenar los datos en Big Data en el mismo servidor, tratar de distribuirlos en varios servidores, cuantos más mejor, de cara a que si un tercero ajeno consigue acceder a algún clúster, sólo consiga averiguar información de uno de ellos, no de todo el sistema de datos en conjunto, de tal forma que, al estar los datos interrelacionados entre varios servidores, la información será muy difícil de entender, incluso, obteniendo solo una de dichas partes necesarias (Mishra & Singh, 2016).

2.2.4.6 Anonimización:

Mediante la Anonimización obtenemos otra valiosa herramienta para proteger los datos masivos recopilados en Big Data, esta idea consiste en utilizar la perturbación de datos y las técnicas de intercambio de datos para proteger la asociación de individuos e información privada que no debe caer jamás en manos ajenas (X. Zhang et al., 2015).

Algunos autores mantienen que el proceso se puede revertir, dando así un resultado negativo de creación y mantenimiento de privacidad de los individuos (Barocas & Nissenbaum, 2014).

Esta herramienta se encuentra íntimamente ligada a la estrategia de diseño “*Minimise*”, según la cual los consumidores de datos sólo deben pedir la cantidad mínima de datos que se requiere para lograr su objetivo. Normalmente, cuando los consumidores de datos solicitan más datos, crean más riesgo para los propietarios de los mismos. En el ámbito de la detección como un servicio, la negociación entre las partes tendrá que tener lugar con el fin de reducir la cantidad de datos que se intercambian mediante la consideración de riesgos asociados y recompensas (Perera et al., 2015). Por ejemplo, si una empresa de venta de productos de limpieza quiere averiguar patrones de consumo y compra de sus productos, no necesita saber el movimiento diario que hace la persona que lo compra para desplazarse hasta el trabajo.

La anonimización tiene que ser algo más que enmascarar o generalizar ciertos campos, los conjuntos de datos anónimos deben analizarse cuidadosamente lo más individualizadamente posible para determinar si son vulnerables a cualquier tipo de ataque (Sedayao, Bhardwaj, & Gorade, 2014).

CAPÍTULO 3. Estado de la Cuestión

Llegados a este apartado, y con una base teórica expuesta anteriormente que ha permitido obtener un bagaje de conocimientos suficiente para conocer todo lo que rodea a Big Data y su aplicación informática y estadística, se procede ahora a estudiar el nivel de interés que se desarrolla de cara a la doctrina específica del término y su aplicación en la realidad, es decir, el interés popular del tema objeto de estudio.

Por pura lógica, este sector servicios se encuentra en la vanguardia de la técnica, el Big Data se trata de un proceso informático y estadístico muy novedoso, aún en formación y en continua mejora para adaptarse a las necesidades que marca el mercado (la tecnología avanzada exponencialmente año tras año, y las cantidades de datos a procesar avanzan similarmente a dicha capacidad, por ende).

De tal forma que en este apartado se procede al estudio de la evolución del interés general que existe sobre Big Data y Privacidad, utilizando para ello Google Trends.

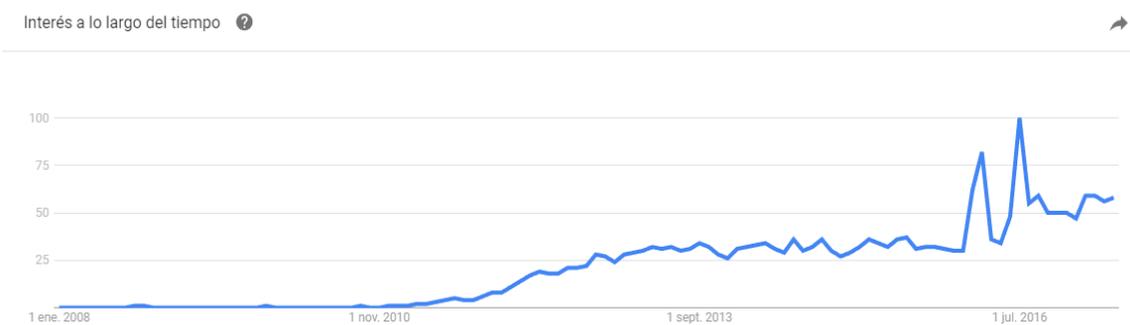
Mediante esta herramienta se puede medir el nivel de búsqueda de un determinado término o conjunto de términos durante periodos de tiempos determinados, escalando los valores relativos de dicho espectro estudiado desde 0 a 100, donde 100 representa el punto más alto alcanzado en los niveles de búsqueda del término estudiado.

Es una herramienta muy potente, permite analizar los niveles de tráfico de los términos en los que estemos interesados, comparar sus niveles de búsqueda con otros, analizar el comportamiento de los usuarios que están interesados en él y las consecuencias que tuviesen las acciones de marketing o empresariales que se llevasen a cabo (mediante el estudio de modificaciones de fechas) e incluso sectorizar geográficamente los niveles de búsqueda.

Por todo ello resulta indispensable utilizar una herramienta tan eficaz como esta para analizar el nivel de interés sobre el vocablo concreto, si goza de buena salud y las variables acaecidas durante el tiempo de cara al estudio e interés de la figura de Big Data.

En nuestro estudio, se comienza con una búsqueda general del término “Big Data”, en el departamento de “búsqueda de noticias” de esta herramienta comentada anteriormente, desde el 1 de enero de 2008 hasta el día 31 de Mayo de 2017, dando como resultado la siguiente gráfica de datos, que procedemos a analizar, para tomar consciencia del nivel de interés en el mismo.

Gráfica 1: Interés a lo largo del tiempo en el concepto



Fuente: <https://trends.google.es/trends/>

En las gráficas de esta herramienta, encontramos que el eje de abscisas representa la variación del tiempo, en nuestro caso concreto, desde enero de 2008 hasta julio de 2016, en cuanto al eje de ordenadas representa el interés de búsqueda en relación con el mayor valor de un gráfico en una región y en un periodo determinados. Un valor de 100 indica la popularidad máxima de un término, mientras que 50 y 0 indican una popularidad que es la mitad o inferior al 1%, respectivamente, en relación al mayor valor. La línea azul remarcada nos especifica el interés determinado relacionado con Big Data.

Como se puede observar el término pasaba bastante inadvertido hasta finales de 2010, a partir de entonces comienza una ascensión paulatina y progresiva, aunque lenta, obteniendo su primer pico de popularidad en octubre de 2012.

A partir de entonces comienza un periodo convulso, de unos 3 años y medio, en los que la popularidad del término cae y asciende consecutivamente, pero con una progresión al alza continua, probablemente por el progreso y desarrollo de investigaciones e interés general en el asunto que ocupa el estudio.

En enero de 2016 se produce con un fenómeno curioso: en los dos siguientes meses a esta fecha el término gozaría de una popularidad casi el 200 % más que anteriormente, no se vislumbra el factor que motiva dicha circunstancia, puesto que después cae en un pico de depresión (relativo, puesto que como se observa vuelve a la media de los cuatro años anteriores) de 3 meses, hasta junio de 2016, donde inexplicablemente se produce un ascenso inverosímil en la popularidad del término, de tal modo que en este periodo encontramos la zona de popularidad más alta del término.

Después de dicho suceso la popularidad decrece hasta la mitad del indicador de interés medido por Google Trends en el momento más popular (es decir, un 50 % menos de búsquedas en cómputo global), y aunque encontramos fluctuaciones en la demanda y búsqueda del término, se puede decir que hasta nuestros días la progresión, aunque lenta, es inexorablemente positiva.

Gráfica 2: Interés por región

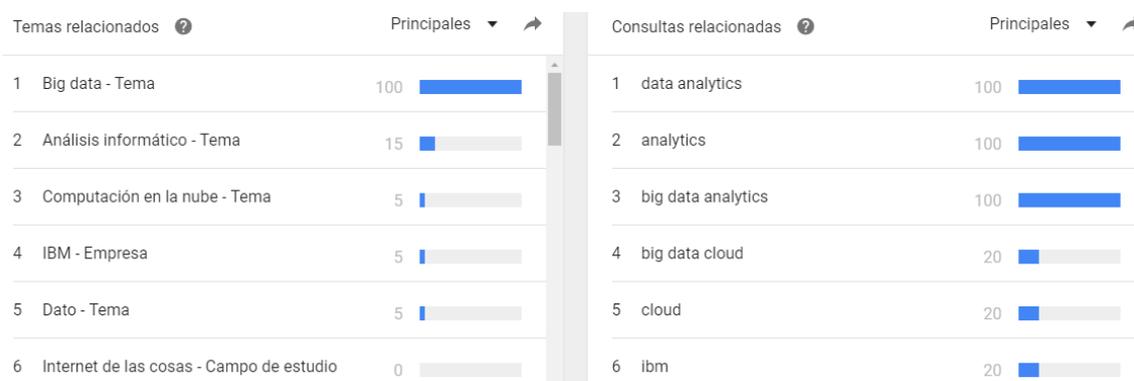


Fuente: <https://trends.google.es/trends/>

En la siguiente imagen se muestra el interés por el término dividido por regiones, en la parte izquierda podemos encontrar los países ilustrados gráficamente y resaltados por colores, en la parte derecha encontramos un ranking elaborado mediante cifras, los valores reflejados se calculan en una escala del 0 al 100, en la que 100 corresponde a la ubicación con mayor frecuencia de búsquedas en proporción al total de búsquedas realizadas en esa ubicación, mientras que 50 y 0 corresponden a las ubicaciones donde la popularidad del término es la mitad o inferior al 1%, respectivamente, en relación a la ubicación con el valor más alto. Encontrándose en los primeros puestos Estados Unidos, Canadá e India, y en un escalafón claramente inferior los países europeos Francia, Inglaterra y España.

Como se observa, este interés se verá fielmente reflejado en el análisis de los datos que se reflejan más adelante este estudio, por ahora puede servir como indicativo interesante de los países que más interesados están en el tema.

Gráfica 3: Keywords más vinculadas



Fuente: <https://trends.google.es/trends/>

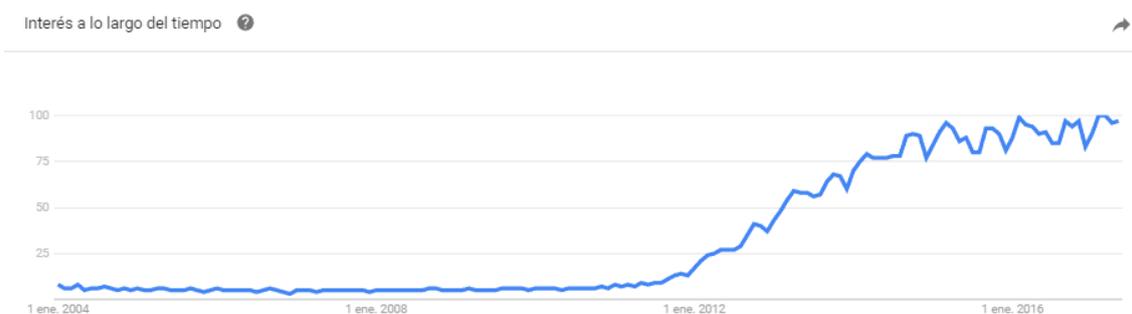
En esta imagen se observan los temas y consultas más relacionadas con Big Data, como es lógico, departamentos de conocimiento como pueden ser el análisis informático o la computación online en la nube están primordialmente relacionados con el Big Data, llama la atención destacadamente que IBM, la conocida multinacional

estadounidense que se encarga de desarrollar tecnología y aportar consultoría especializada en el tema, se encuentre en cuarta posición, incluso por delante de términos como “Dato” o incluso el campo de estudio “Internet of Things”, tan de moda en los últimos tiempos.

En cuanto a las consultas, keywords como “data analytics”, “Big Data cloud” o sencillamente “cloud”, se encuentran estrechamente vinculadas con el tema que nos atañe, como además se puede comparar y desarrollar en el capítulo quinto de este estudio titulado “Estudio Bibliométrico”.

En las siguientes gráficas del estudio se establece un nuevo criterio de búsqueda, una modificación simple pero efectiva, sobre todo de cara a la comparativa, utilizando el mismo término pero esta vez analizando el apartado “búsqueda web”, es decir, el departamento anexo que nos presenta como opción la herramienta Google Trends.

Gráfica 4: Interés a lo largo del tiempo, “búsqueda web”



Fuente: <https://trends.google.es/trends/>

Como se observa se sigue manteniendo la misma línea de progreso en líneas generales, a partir de 2012 comienza un ascenso de la popularidad del término considerable, teniendo picos depresivos de audiencia, pero en general manteniendo un cupo de visualizaciones y búsqueda tendente al progreso y a aumentar su número progresivamente.

Gráfica 5: Interés por región en “búsquedas web”

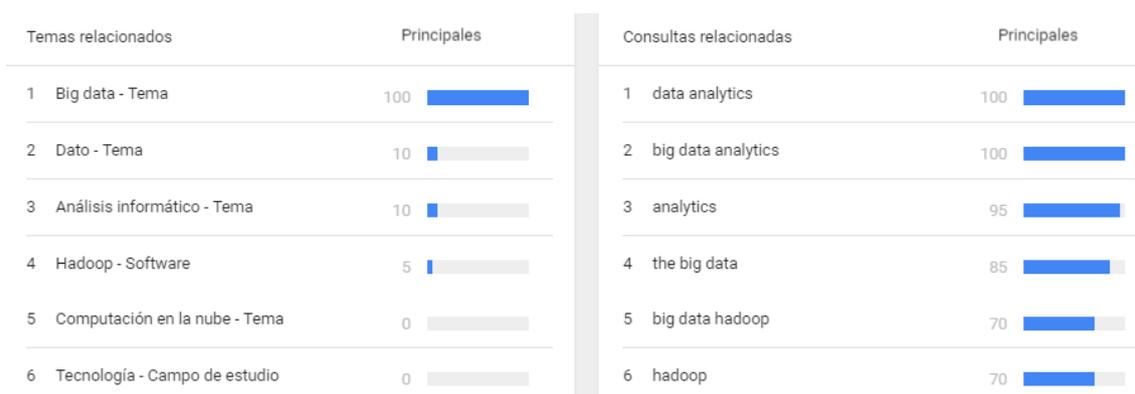


Fuente: <https://trends.google.es/trends/>

En el apartado de interés por región se comienzan a ver diferencias, los países más interesados dejan de ser Estados Unidos y las principales potencias Europeas para dar paso a los principales países asiáticos, especial referencia a India, que es el único país que se mantiene en el “Top Five” tanto en la clasificación anterior como en la actual, lo cual denota un enorme interés por parte del país y le aúpa a la posición número 1 del ranking general (teniendo en cuenta ambas clasificaciones, se entiende).

En cuanto al resto de países, resulta llamativo con el análisis bibliográfico que más tarde se explicará con atención (apartado quinto del estudio), pero la mayoría de esos países asiáticos que figuran en la lista no tienen un reflejo fiel de dicho interés en la mayoría de artículos encontrados y estudiados, quizá sea porque existe mucho interés social en el tema, pero no cuentan con empresas e industrias de renombre que puedan llevar a cabo la financiación de investigaciones sobre el tema objeto de estudio.

Gráfica 6: Keywords más vinculadas a “búsquedas web”



Fuente: <https://trends.google.es/trends/>

Como se puedes apreciar, los términos mas relacionados con esta búsqueda son “Big Data”, “Dato”, “análisis informático” o “Hadoop”, y en cuanto a las consultas relacionadas se encuentran Big data, analytics y hadoop, asi como la combinación que forman las mismas palabras entre sí.

Por último, este estudio se centrará en la utilización de esta herramienta, pero en el estudio de las palabras más específicas en relación con el tema objeto de estudio “Big Data Privacy”.

Gráfica 7: Interés a lo largo del tiempo de “Big Data Privacy”

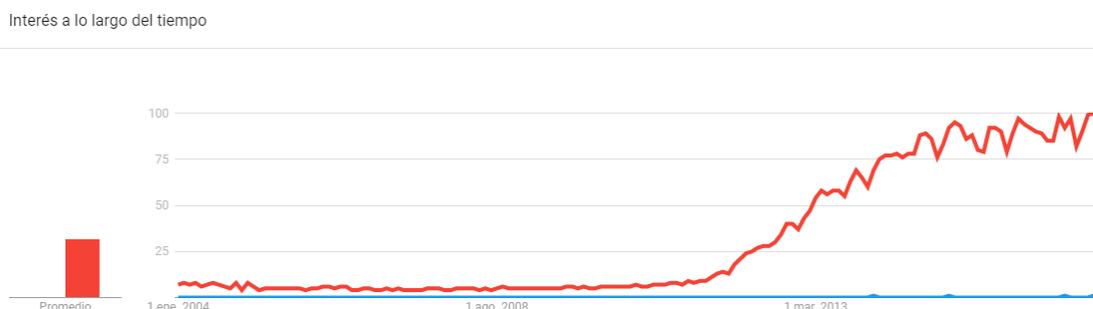


Fuente: <https://trends.google.es/trends/>

En este caso ocurre algo bastante extraño, podemos observar un gran interés en el tema hace más de una década, concretamente en 2004, seguido de procesos de profunda depresión en el interés del término, que van simultaneándose con picos de interés, concretamente en mayo de 2005, marzo, julio y agosto de 2006; desde 2007 hasta 2013 encontramos un desinterés completo sobre el tema, hasta finales de 2013 donde encontramos un ascenso paulatino del interés sobre el término, excepto en dos fechas, verano de 2014 y verano de 2015, el resto de fechas sigue un orden in crescendo, denotando un mayor interés cada día más completo en la Privacidad que se maneja en el Big Data.

La siguiente gráfica muestra una comparativa entre el interés de búsqueda de las palabras “Big data” en color rojo, frente a “Big data Privacy”, en azul.

Gráfica 8: Interés a lo largo del tiempo de “Big Data Privacy”



Fuente: <https://trends.google.es/trends/>

Como se puede observar, la comparación entre ambas no deja lugar a dudas, las búsquedas web del término genérico (es decir, sin el acompañante Privacidad) dan como resultado muchísimas más visitas e interés por parte de las personas de todo el mundo. Se pueden observar pequeños incrementos en la línea azul, procedente de la Privacidad aplicada a Big Data, pero es totalmente insignificante frente al interés general que aporta el término normal en la red de redes.

A modo de conclusión, se observa cómo tanto en el término general (Big Data) como en el término específico (Big Data Privacy) hay dos claras tendencias, una primera en la que el término se encuentra en lo que se puede llamar “estado de gestación”, que en todas las gráficas observadas rondan entre finales de 2012 y comienzos de 2013, justo cuando, como se analizará más adelante, comienza a existir más doctrina especializada en el tema, además comienza la segunda tendencia, el que podríamos llamar “estado de explotación”, es decir, el periodo desde el cual, a pesar de que existan pequeños periodos de depresión, el interés sobre el tema se ha mantenido bastante alto y en continuo aumento.

En cuanto a la cantidad de visitas de un término a otro tan dispares, se puede llegar a la conclusión de que el término Big Data Privacy no es buscado como un tema de ocio o interés popular, de ahí que existan con unas diferencias tan contrastadas en los resultados de búsqueda de ambos términos, seguramente el término más específico quede relegado a estudios o revisiones bibliográficas como las que trata de llevar a cabo este trabajo, y por ende los resultados en navegadores resultan en cifras mucho más humildes.

Para finalizar, resulta bastante destacable en mi opinión personal cómo varía el interés por región si mediante la herramienta Google Trends se aplican búsquedas de noticias sobre el tema (dando como resultado un interés mayúsculo en estados unidos y naciones europeas) y cambiando drásticamente si aplicamos búsqueda web, resultando como tal India y países orientales (es decir, ámbitos completamente diferentes).

En general se observa un Estado de la Cuestión sobre Big Data con muy buena salud y pronóstico futuro, el interés tanto popular como académico sobre el tema sigue en auge, aunque en años pretéritos haya sufrido ciertas recaídas en algunas de las gráficas observadas, pero siempre manteniendo un progreso al alza más o menos progresivo, lo cual denota un interés que, como se plasmará más adelante en el análisis bibliográfico llevado a cabo, tiene su representación en un mayor número de estudios y artículos académicos, así como en la calidad de los mismos.

CAPÍTULO 4. Otros estudios bibliométricos sobre Big Data y Tecnologías de la Información. La importancia de estudios análogos

La importancia capital del tema de estudio de este trabajo queda consensuada cuando se analiza a la cantidad de estudios y revisiones bibliométricas existentes sobre el tema en todo del mundo y de los estudios llevados a cabo por diferentes autores.

Así se encuentran estudios como el del autor López-Borrull & Canals (2013), donde se aborda un estudio bibliográfico, tanto a nivel cualitativo como a nivel bibliométrico, llevado a cabo utilizando la base de datos Web of Knowledge sobre Big Data (además de Open Science y e-Science), en el periodo de tiempo entre 2000 y 2013 (año en el que se publicó el artículo).

El artículo recoge una gráfica parecida a la que aporta nuestro estudio de artículos publicados sobre el tema clasificados por su año de publicación, dando una referencia o conclusión similar a la que se alcanza en nuestro apartado análisis de datos, es decir, una eclosión de artículos sobre el tema exponencial en los años 2012 y 2013, y refrendando la idea que defiende nuestro análisis bibliométrico de que en esas fechas la doctrina Big Data comenzó a ser lo suficientemente popular e interesante como para que la doctrina comenzase a producir conocimiento e investigación en grandes cantidades.

Se hace también hincapié en los numerosos desafíos de manejo de datos, privacidad y eficiencia al que deberán exponerse los trabajadores que manejen las herramientas de Big Data (la dualidad que se analiza posteriormente en las conclusiones de este trabajo), aportándole un gran valor en el ámbito aplicativo, tanto como las herramientas tecnológicas de las que se dispongan ofrezcan. Plantea la posibilidad y defiende la idea de que Big Data ayude a que grupos de investigadores de diferentes países compartan fácilmente y comparen datos de temas comunes de investigación, analizando metodologías, indicadores y parámetros manejados en cada investigación, pero que nutran de forma colectiva a todos los autores interesados en el tema, algo así como la creación de una conciencia colectiva sobre el tema objeto de estudio.

Se encuentra también en dicho estudio, la idea de Big Data con la filosofía Open (acceso al máximo número de conocimientos de forma libre y gratuita, como método para alcanzar el proceso global) preguntándose si las herramientas aplicativos de Big Data serán capaces de digerir, gestionar y tratar toda esa ingente cantidad de datos necesaria.

Como conclusión alcanza que cualquier proyecto de Big Data, debe estar profesionalmente diseñado y metodológicamente cuidado para poder crear conocimiento y que éste obtenga una continuidad y capacidad suficientes para llegar a ser eficaz en la finalidad que persigue (López-Borrull & Canals, 2013).

Otros estudios se centran en combatir el fraude contable mediante técnicas aplicativas de Big Data, como puede ser el Data Mining, en el artículo del autor Alexander Báez Hernández (Revista Publicando, 2015); artículo en el que se lleva a cabo una completa revisión bibliométrica sobre los términos “data mining” y “fraud”.

La revisión de la literatura llevada a cabo en dicho estudio permitió comprobar que la aplicación de las técnicas de la Minería de datos puede ser de gran eficacia para la detección de los distintos tipos de fraude contable, de hecho, tratándose de una línea de investigación con gran bagaje de cara al futuro. En sus conclusiones defiende que la técnica de Data Mining es la herramienta idónea para manejar todos los ámbitos necesarios en relación con los datos contables y su estudio correcto y eficaz para la detección de fraudes en los mismos (Revista Publicando, 2015).

En el ámbito de las Tecnologías de la Comunicación, altamente ligado al ámbito aplicativo del Big Data y su Privacidad en las herramientas que lo involucran, encontramos otros estudios como el del autor Facundo Froment (2017), donde se realiza un estudio de como se ha modificado las relaciones entre los profesores y docentes de las universidades y los alumnos que llevan a cabo las formaciones de los mismos en los últimos tiempos, con el auge de las redes sociales y las demás herramientas que se han puesto a nuestra disposición, tales como Instagram, Facebook o Twitter (Froment, 2017).

El procedimiento utilizado en la metodología de dicho estudio es bastante similar a la de este estudio llevado a cabo. Primero procede el autor a definir los determinados agentes de búsqueda y booleanos utilizados para dar con los artículos y demás material académico estudiados, en este caso concreto más tarde el autor utilizó un sistema cerrado de categorías de cara a clasificar y estudiar los artículos; este apartado no es exactamente igual, puesto que dicho proceso es altamente personalizable dependiendo de los gustos del autor, pero al fin y al cabo lo que se necesita es alcanzar un método por el cual se planifiquen y gestionen los artículos adecuadamente de cara a su estudio, que es lo que se persigue en el fin último, y dichos métodos son comparables entre ambos estudios, sobre todo las características perseguidas en los criterios de búsquedas en las bases de datos de Internet.

Tras el estudio, presenta una serie de datos para refrendar su estudio bibliométrico tales como el tipo de documentos al que pertenecía el estudio (artículos de revista, comunicaciones de congresos, tesis doctorales, etc...), una lista de los autores más productivos, una clasificación de la cantidad de documentos firmados por cada autor (destacando los más predominantes), una clasificación dependiendo de si los autores son mujeres u hombres, una clasificación dependiendo del año de publicación de los artículos, incluso clasifica las revistas más productivas en crear y publicar artículos, los artículos con referencias más empleadas y la afiliación institucional de cada uno de los artículos (clasificación de universidades); como podemos ver, muchas de esas clasificaciones se encuentran en el estudio bibliométrico llevado a cabo en este texto, y siguen líneas muy comunes.

En el último tramo del artículo, se presenta un apartado titulado “Discusión”, que viene a desarrollar el objetivo que se pretendía alcanzar con el estudio bibliométrico llevado a cabo, su principal conclusión radica en la escasa doctrina existente en el tema y el amplio abanico de posibilidades de extensión de conocimiento que aún queda por desarrollar sobre el tema; este apartado también cuenta con un homólogo en cada una de las partes (tanto de desarrollo teórico como de conclusiones) en el apartado final del texto que tiene entre manos, donde se desarrollan los objetivos y conclusiones alcanzados sobre el estado de la protección de la Privacidad en el desarrollo y aplicación de Big Data.

Estudiando otro tipo de artículos, esta vez más enfocados hacia la administración pública y gubernamental de los países, se encuentra el artículo de Laura Alcaide Muñoz, et al. (2013) de la mano de tres profesores de la Universidad de Granada. En este documento se encuentra un estudio profundo sobre la introducción de las nuevas tecnologías en las reformas administrativas del sector público, siendo muchas de ellas referentes a la inclusión de nuevas tecnologías, por ello se convierte en un estudio/examen sobre el estado en el que se encuentra la investigación en la implementación de las tecnologías de la información y la comunicación en las administraciones públicas; para ello, analiza las revistas de más prestigio e impacto internacional que se encuentran incluidas dentro de las categorías “Administración Pública” y “Ciencias de la Información”.

Tras ello presenta una serie de tablas informativas a modo clasificatorio, tales como “artículos publicados en cada revista” o “distribución cronológica de temas de investigación en E-Gobierno”, aportando datos de gran valor que son muy útiles como fuente en futuros estudios sobre el tema. En el apartado referente a las conclusiones del trabajo, se analizan ideas interesantes, como una de ellas en la que aporta que el campo de investigación e-Gobierno es muy complejo, capaz de reunir diferentes campos de estudios en uno solo, aportando metodologías propias en cada área y técnicas individuales, que a la vez crean problemas para conseguir una maduración concreta en el tema específico que les atañe.

Otra de las conclusiones alcanzadas es que mientras estos estudios son muy comunes en países anglosajones (para valorar la efectividad de los cambios realizados en sus administraciones públicas) no lo son tanto en países cuyas economías son menos favorecidas o desarrolladas, de los cuales opina que habría que estudiar y tratar las decisiones de los responsables políticos en la materia que les ha llevado a implementar o no diferentes medidas, y cómo aplicarlas efectivamente tras los casos conocidos en otros países. Aporta también que los investigadores que más han aportado a esta rama de investigación son aquellos ligados a departamentos de administración pública, ciencias públicas y políticas, y de gestión y organización de empresas.

Por último, ayuda e insta a seguir investigando sobre el tema en cuestión tanto de conocer el estado de la participación ciudadana, de la e-Democracia y de la e-Gobernanza, delimitando y estudiando los factores que podrían favorecer la implantación y desarrollo de esos aspectos, emplazando tales temáticas como un campo de investigación bastante interesante para futuros proyectos y estudios (Muñoz, Pedro, & Bolívar, 2013).

Se pueden encontrar otros artículos, como el de María Sáinz, et al. (2013). Como dos grandes conclusiones en este estudio se alcanzan, la primera de ellas es que el uso de herramientas para compartir conocimiento (entre ellas Big Data), encuentra un ámbito aplicativo ideal en el periodismo de datos debido a las aplicaciones comunicativas que tienen lugar, dando lugar a numerosas posibilidades para los profesionales del periodismo, para acceder y procesar grandes volúmenes de datos. La segunda de las conclusiones es que los cinco casos de estudio que se manejan en el artículo, confirman el interés creciente en los datos e historias que se pueden desvelar a través de análisis de Big Data (Sáinz, Iñigo, Raquel, & Sanz, 2013).

En el artículo llevado a cabo por Alberto Calero (Revista Telos, 2013), como estudio complejo de los nuevos métodos de análisis de datos, denominándolos “Nueva Ciencia de las Redes”, se encuentra completamente enfocado y dirigido al ámbito aplicativo que nos interesa, y lleva a cabo un estudio en diferentes facetas de aplicación, tales como seguridad, marketing, mercados financieros, sanidad, telecomunicaciones, etc... Concluyendo con la idea de que esta nueva ciencia ha alcanzado un nivel de notoriedad dentro del campo de Big Data bastante destacable, permitiendo extraer relevancia de forma rápida de las cantidades de datos que se maneja.

Asimismo, en el artículo llevado a cabo por la autora Soraya Paniagua (Revista Telos, 2013), se adentra en un estudio bastante específico relacionado con la forma de obtener los datos que se manejan en Big Data y dónde se crean los mismos. Sostiene que el bajo consumo, el abaratamiento de los dispositivos y las comunicaciones inalámbricas que se tienen a disposición hoy día, facilitan el tratamiento de grandes datos en la sociedad.

Otorga ejemplos de cómo los sensores de datos facilitan actualmente la vida en muchas facetas de la sociedad. Determina que claramente la mayor fuente de datos e información creada diariamente se encuentra en Internet, explicando cómo Google, para manejar la gran cantidad de datos que los usuarios creaban en el uso diario de las aplicaciones que habían puesto a su disposición, se vieron obligados a crear la herramienta Hadoop, para tratar dicha información indexada, herramienta de la que hemos hablado en nuestro estudio. Termina este estudio con la conclusión firme de que actualmente estamos viviendo una nueva revolución, pero a diferencia con décadas anteriores (revolución industrial, informática, etc...), la que vivimos actualmente es referente a los datos y la información.

Estos tres últimos artículos estudiados y analizados se encuentran en un compendio o dossier de estudios relacionados con el ámbito y el paradigma investigador que rodea al Big Data, publicados todos ellos en el año 2013 bajo el nombre de “Telos, Cuadernos de Comunicación”, con colaboración de Fundación Telefónica (Sáinz, Iñigo, Raquel, & Sanz, 2013), en el cual se reúnen autores intelectuales con clara inclinación e interés sobre el tema.

Lo cual ayuda a entender, como se puede observar, que el tema de interés es predominante en la doctrina científica actual, que no solo moviliza a autores independientes, sino que incluso en España es tendente a crear asociaciones y dossieres como el estudiado con múltiples grupos de interés sobre el tema, ya no solo para dar a conocer y expandir la utilización de estas herramientas, si no para profundizar, analizar y extraer conclusiones y nuevas visiones sobre los temas más candentes, los desafíos y oportunidades que se pueden abordar en futuras investigaciones, los procesos y mejoras que aún quedan por plantear y crear, de cara a que el futuro de la investigación en Big Data y procesamiento de datos sea tan productivo y eficaz como lo es en la actualidad.

Todos los estudios analizados aportan datos interesantes en diferentes ámbitos de la materia, ya sea de Big Data o de Tecnologías de la Información, incluso muchos de ellos aportan proyectos de estudios futuros y nuevas líneas de investigación, lo cual permite dotar de perspectiva a la hora de conocer la gran importancia que tienen las investigaciones desarrolladas sobre el tema actualmente, ya no solo por la gran cantidad de artículos que encontrados sobre Big Data y sus aplicaciones, sino de la calidad, variedad y cantidad de información que aportan y desarrollan todos ellos.

CAPÍTULO 5. Análisis Bibliométrico

5.1. Metodología

La metodología llevada a cabo para realizar el estudio bibliográfico de este Trabajo Fin de Máster se centra en analizar las búsquedas de artículos doctrinales, ponencias, libros y comunicaciones de congresos de prestigio (en pos de conseguir la máxima calidad y cantidad de contenido posible) en 3 bases de datos online que se procede a explicar con una pequeña referencia de cada uno de ellos.

En cuanto a la acotación temporal de la búsqueda, se realizó la búsqueda entre el año 2000 hasta la actualidad, en la primera búsqueda bibliográfica se incluyeron 3 artículos del año 2017 que no aparecen en graficas como “número de artículos por año”, debido a que desvirtuaban el significado de la misma (reflejaba un número de artículos mucho menor que años anteriores), pero a efectos de datos se han tenido en cuenta de cara a otras gráficas, por dicho motivo el número de artículos resultante al sumar los datos de unas u otras puede variar. Para dar con el mayor número de artículos e información sobre el tema, se llevaron a cabo una serie de búsquedas bibliográficas (siguiendo unas pautas establecidas y unos criterios de calidad, que más adelante se procede a explicar detalladamente) para seleccionar y analizar aquellos artículos científicos que efectivamente trataran sobre Big Data y la privacidad de los datos manejados en el mismo. La búsqueda llevada a cabo se centró en las principales bases de datos de referencias bibliográficas y citas de publicaciones periódicas a nivel mundial, como artículos de revistas científicas, libros y otros tipos de material impreso.

Estas bases de datos son:

1) Web of Science, propiedad de *Thomson Reuters*, se trata de un servicio en línea de información científica, se encuentra integrado en “ISI Web of Knowledge”. Se trata de un recurso web con el que se accede a un conjunto de bases de datos en las que, mediante criterios de búsqueda, se analizan artículos de revistas científicas, libros y todo tipo de material impreso de todas las disciplinas de conocimiento académico.

2) Scopus, propiedad de *Elsevier* (Ministerio de Economía y Competitividad, 2015) se trata de la mayor base de datos de resúmenes hasta ahora en el mundo, en su haber goza de más de 20.000 publicaciones, procedentes de más de 5.000 editoriales internacionales, dota a los estudiantes e investigadores del acceso a más de 28 millones de resúmenes y su contenido es actualizado semanalmente. Ofrece una herramienta muy útil y potente de cara a los investigadores y estudiantes de todas las áreas, poniendo a su disposición el acceso a textos completos a los artículos de investigación estudiados.

3) FAMA US, propiedad de la Universidad de Sevilla, se trata del catálogo de la Biblioteca de la misma entidad, el cual goza de un excelente buscador avanzado de artículos, libros, revistas y demás material académico y docente, no es una herramienta tan potente y con un catálogo tan ingente como las dos anteriores, pero permitió realizar búsquedas de artículos sobre la materia para servir de apoyo al grueso de los artículos encontrados en las dos bases de datos anteriores.

Los parámetros de búsqueda en los apartados “búsqueda avanzada” de cada uno de estos buscadores que se llevaron a cabo son los que se exponen a continuación:

Web Of Science: Parámetro de búsqueda, TI=(big* data* privacy*), hace un total de 141 resultados.

Scopus: Parámetro de búsqueda, TITLE(big) (data) (privacy), hace un total de 1.494 resultados.

Si a su vez dentro de esta búsqueda acotamos por “Business Management and Accountig” (entendido como el área más similar al estudio del Máster y de la rama del estudio), la búsqueda desciende a 121 resultados.

A su vez filtro por los keywords o etiquetas “Big Data” y “Data Privacy”, los resultados se restringen a 78.

FAMA US: Parámetro de búsqueda, =(big) (data) (privacy), hace un total de 254.026 resultados. Si a su vez se acota por los resultados en español, se obtiene un montante final de 68 resultados, que podrían servir como apoyo a las dos búsquedas anteriores de documentos.

El montante final de resultados de las 3 bases de datos relacionales **sumarían 287 documentos**, los cuales fueron sometidos al siguiente proceso de depuración y tratamiento:

- *Fase 1:* Fueron analizados uno por uno los artículos a través del volcado de los datos de los mismos en el programa informático Mendeley, el análisis del “*Abstract*” o el resumen de los artículos permitió realizar una criba de calidad, ya sea por artículos que no se adaptaban al tema o por documentos que aportaban información excesivamente técnica perteneciente a otros sectores de estudio, se alcanzó la cifra definitiva de artículos que se estudian en este estudio bibliográfico, eliminando los artículos que no se asemejaban al tema de interés. Los artículos de mayor interés quedaban marcados mediante el símbolo de favorito, y los que seguían remarcando alguna duda en lo referente a si me terminarían resultando útiles para la investigación quedaron marcados con un punto rojo.

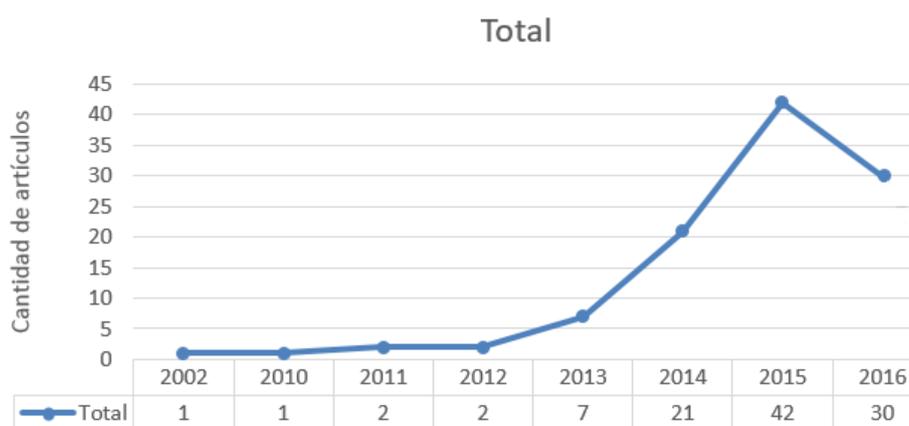
- *Fase 2:* Una vez con el montante final de artículos y libros, se procedió a la búsqueda y descarga del archivo de texto de los mismos para adjuntarlo a su referencia en el programa informático comentado anteriormente, de cara a obtener la máxima información del mismo: autores, año de publicación, etc... y a tener disponible su desarrollo de contenido para el estudio posterior del mismo (el cuerpo de muchos de esos artículos resultó fuente de información la cual nutrió el cuerpo teórico del capítulo segundo de este estudio).
- *Fase 3:* Una vez que se dispuso de la gran mayoría de los artículos de manera completa, se llevó a cabo la selección final así como la eliminación de duplicados, puesto que algunos se encontraron tanto en Scopus como en Web of Science, en la selección final se procedió a una lectura más en profundidad de aquellos artículos que comprendían algún tipo de duda (marcados anteriormente con el punto rojo) para contemplarlos en el análisis bibliográfico. El montante final de artículos y libros a estudiar y tratar fue de **106** (109 teniendo en cuenta los artículos residuales de 2017 comentados anteriormente).
- *Fase 4:* Exportación de datos desde Mendeley al programa Excel del paquete Microsoft Office, donde se procedió a grabar los datos (autores, año de publicación, entidad productora, etc...) de todos los artículos con los que más tarde se moldearían las tablas y gráficos de los cuales sacar conclusiones de la investigación en la materia. Otras investigaciones similares otorgaron una guía sobre la que obtener unos parámetros y consejos para llevar a cabo el estudio y trata de datos de todos los artículos que se había encontrado (Congreso et al., 2016).
- *Fase 5:* Creación de las tablas y gráficos con los que sacar conclusiones de la investigación en la materia gracias a la base de datos creada anteriormente, sin la cual hubiese sido imposible plasmar la investigación completa y el análisis de datos que se expone en el siguiente capítulo del estudio.

5.2. Análisis de Datos

En este apartado de la investigación se plasmarán los resultados obtenidos tras el registro y moldeamiento de los indicadores principales de los 106 artículos seleccionados para la revisión bibliográfica sobre Big Data y Privacidad. Estudiaremos el tema desde diferentes puntos de vista y comparando diferentes indicadores.

5.2.1. Número de artículos por año

Gráfica 9: Número de artículos por año



Fuente: Elaboración propia

Como podemos observar en la gráfica que acompaña a este apartado (Gráfica 9), la preocupación y la investigación desarrollada alrededor de la Privacidad de los datos en Big Data ha ido aumentando desde el año 2000 aproximadamente.

De la gráfica también se puede sacar la conclusión de que Big Data es una disciplina muy moderna y actual, la cual aún no está completamente enraizada en la investigación internacional, puesto que lleva muy pocos años en desarrollo. El primer artículo encontrado sobre el tema tiene quince años de antigüedad (Evfimievski, 2002).

Se evidencia un incremento significativo en el número de artículos, especialmente en los años 2012 a 2013 y las dos siguientes etapas anuales, donde el número de artículos se llega a duplicar (o incluso más) en un periodo muy breve de tiempo. Esto no es más que el reflejo de la sociedad de la información en la que vivimos, el concepto se mantuvo en estado de “hibernación” en un comienzo, probablemente cuando el concepto aun ni si quiera fuese material, no existiesen herramientas populares de uso y aplicación de Big Data o no estuviese popularizado su uso, solo unos contados especialistas en la materia apostaban por el estudio de esta doctrina.

Este periodo se podría enmarcar desde el año 2002 hasta el 2012, 10 años en total, periodo en el cual solo se encuentran 6 artículos sobre el tema, si se compara con la cantidad de artículos encontrados en años posteriores, la cantidad de artículos se multiplica exponencialmente.

Es decir, en dicha fase es probable que el término Big Data estuviese estudiándose, y como en todo concepto, al comienzo, la mayoría de los estudios se centrarían en temas más prácticos, como puede ser la aplicabilidad de las aplicaciones y herramientas informáticas que hacen posible el manejo de enormes cantidades de datos en Big Data, también habría muchos artículos haciendo ver las ventajas de aplicar esta herramienta y no otras aplicaciones más convencionales que dificultarían más trabajar con grandes datos (Salleh & Janczewski, 2016).

Pero a cuestiones como la Privacidad, que se puede decir resultan un freno hacia el desarrollo práctico y aplicativo del concepto, quizás se tardara más en llegar, manteniendo ese letargo los primeros diez años de existencia del término, y encontrándose más tarde cuando comenzaron los primeros problemas de intimidad de datos de los usuarios y las personas de las cuales se manejaban datos en Big Data.

De ahí que se evidencie un pequeño lapso de tiempo de 3 años, entre 2013 y 2015, con el contraste enorme de 70 artículos y material académico sobre el tema en dichos años, que resulta una cifra aplastante si se compara con los 6 artículos de toda la década anterior de investigación.

Mención aparte merece el año 2016, donde se puede observar una disminución en el número de artículos encontrados sobre el tema que incumbe, frenando así el aumento gradual que se venía produciendo desde unos pocos años atrás, para un buen observador será fácil encontrar que el número de artículos sigue siendo bastante importante, para ser un tema tan específico y sigue siendo superior al año 2014, lo cual significa una enorme vitalidad en el desarrollo doctrinal del tema.

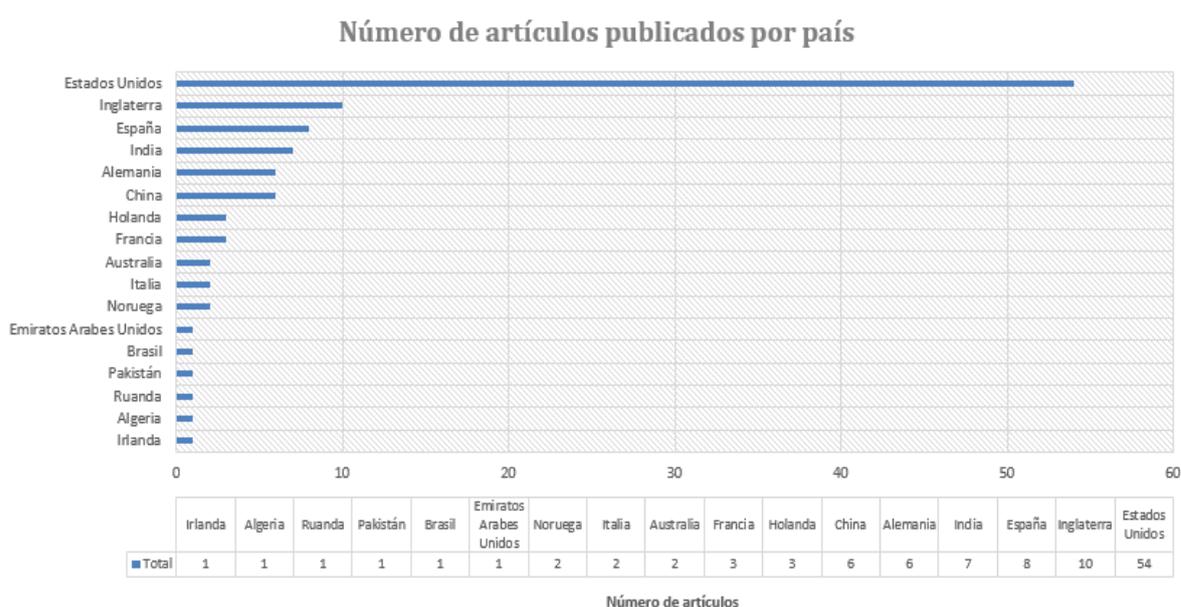
La cifra del año que ocupa es meramente anecdótica, solo tres artículos (no reflejados en el gráfico para no desvirtuar el mismo), esto es debido a que el estudio bibliográfico se realizó a comienzos del mismo año, por lo cual es imposible plasmar una cifra verdaderamente fidedigna y representativa de los artículos que finalmente se publicaran sobre la Privacidad de los datos en Big Data, por motivos evidentes, siendo la cifra resultante a búsqueda día 23 de Marzo de 2017, 3 artículos sobre el tema. Una segunda búsqueda a fecha 17 de Julio de 2017 nos puede ayudar para entender la conclusión alcanzada en este apartado, siendo que en pocos meses la cifra de 3 artículos ha aumentado en los siguientes resultados:

- Mediante la plataforma web **Scopus**, siguiendo el parámetro de búsqueda comentado anteriormente, se han encontrado 35 artículos limitándolo a su vez en 2017.

Colocándose en una posición pareja a los años 2015 y 2016, teniendo en cuenta que hemos sobrepasado un mes del equinoccio del año presente, se puede aventurar que este es probablemente el año que más artículos se van a publicar sobre el tema que nos ocupa hasta la fecha, y reforzando la idea de que el interés investigador sobre la Privacidad en Big Data no hace sino aumentar año tras año.

5.2.2. Número de artículos por país

Gráfica 10: Número de artículos por país



Fuente: Elaboración propia

Como se puede observar, se encuentra un mayor número de artículos y demás material académico sobre la privacidad de los datos en las herramientas utilizadas por Big Data en Estados Unidos.

Se pueden establecer tres niveles de aportación al tema que interesa, en el primer nivel se encuentran en clara ventaja y en solitaria situación a Estados Unidos con 54 artículos, referente mundial en el género, la comparación frente a cualquier otro país del mundo puede llegar incluso a ser irrisoria.

En el segundo nivel encontramos cinco países, Inglaterra, en segunda posición con 10 artículos estudiando el tema, como se puede observar no se trata ni si quiera de una quinta parte de lo que ofrece el país con mayor número de artículos; seguido muy de cerca por España e India (con 8 y 7 artículos respectivamente), es motivo de orgullo que el país en el que resido se encuentra en tercera posición en el ranking mundial, puesto que no se financia demasiado en investigación y normalmente las herramientas

y recursos para ello son más difíciles de encontrar y utilizar, pero es motivo de alegría, puesto que demuestra una clara apuesta por el futuro y es característico de una sociedad que busca el futuro informativo y tecnológico de la misma forma que lo hacen otros países desarrollados y a la vanguardia de la técnica.

Incluidos en el segundo nivel se encuentran Alemania y China, con 6 artículos otorgados en el tema que se estudia, cerrarían el segundo escalafón de países más importantes en su colaboración con la Privacidad en Big Data.

En el tercer y último escalón se encuentran todos los demás países que han aportado de forma breve o concisa, teniendo entre una o 3 publicaciones todos ellos, que son los siguientes: Algeria, Australia, Brasil, Emiratos Árabes Unidos, Francia, Holanda, Irlanda, Italia, Noruega, Pakistán o Ruanda. Estos dos últimos llaman la atención en la lista, puesto que no son potencias mundiales económicas, pero por algún motivo tienen algún tipo de interés en el tema objeto de estudio.

5.2.3. Número de artículos por revista

Tabla 1: Artículos por revista

Revista	Artículos por revista
Institute of Electrical and Electronics Engineers (Recuento de todas las revistas de la tabla)	22
Association for Computing Machinery (recuento de todas las revistas de la tabla)	6
Computer Law & Security Review	5
Procedia Computer Science	3
Big Data Research	2
Institute of Electrical and Electronics Engineers Cloud Computing	2
Institute of Electrical and Electronics Engineers International Congress on Big Data	2
International Journal of Managing Projects in Business	2
Accounting Horizons	1
Association for Computing Machinery Communications of the ACM	1
Association for Computing Machinery International Conference Proceeding Series	1
Association for Computing Machinery Queue	1
Association for Computing Machinery SIGKDD Explorations Newsletter	1
Big Data Management	1
Business Horizons	1
Challenges of Information Management Beyond the Cloud	1
Cloud Security Alliance	1
CODASPY (Conference on Data and Application Security and Privacy)	1
Concurrency and Computation-Praxtice & Experience	1
Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, CSIBIG	1

Digital Investigation	1
EPJ Data Science	1
Government Information Quarterly	1
Information Society	1
Information Sciences	1
Information Systems Design and Intelligent Applications	1
Institute of Electrical and Electronics Engineers Access	1
Institute of Electrical and Electronics Engineers China Summit & International Conference on Signal and Information Processing (Chinasip)	1
Institute of Electrical and Electronics Engineers Conference on Computer Communications Workshops (Infocom WKSHPS)	1
Institute of Electrical and Electronics Engineers Engineering Management Review	1
Institute of Electrical and Electronics Engineers International Conferences on Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/SCALCOM/CBDCOM/IOP/SMARTWORLD)	1
Institute of Electrical and Electronics Engineers International Conference on Computational Intelligence and Communication Technology 2015	1
Institute of Electrical and Electronics Engineers International Conference on Computer Communication and the Internet (ICCCI 2016)	1
Institute of Electrical and Electronics Engineers International Conference on Computing, Communication and Automation (ICCCA)	1
Institute of Electrical and Electronics Engineers International Conference on Electric Utility Deregulation, Restructuring and Power Technologies, DRPT 2015	1
Institute of Electrical and Electronics Engineers International Congress on Big Data	1
Institute of Electrical and Electronics Engineers Long Island Systems, Applications and Technology Conference (LISAT)	1
Institute of Electrical and Electronics Engineers Network	1
Institute of Electrical and Electronics Engineers Proceedings - Institute of Electrical and Electronics Engineers Infocom	1
Institute of Electrical and Electronics Engineers Proceedings, Institute of Electrical and Electronics Engineers International Conference on Big Data	1
Institute of Electrical and Electronics Engineers Symposium on Computers and Communication (ISCC)	1
Institute of Electrical and Electronics Engineers Transactions on Computer	1
Institute of Electrical and Electronics Engineers Transactions on Multimedia	1
Institute of Electrical and Electronics Engineers/Association for Computing Machinery International Conference on Advances in Social Networks Analysis and Mining Asonam	1
Institute of Electrical and Electronics Engineers/Association for Computing Machinery International Conference on Utility and Cloud Computing (UCC)	1
Institute of Electrical and Electronics Engineers/EAI International Conference on Collaborative Computing, CollaborateCom	1
International Business Information Management Association Conference - Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth	1
International Computer Science and Engineering Conference (ICSEC)	1
International Conference for Internet Technology and Secured Transactions (ICITST)	1
International Conference on Big Data and Smart Computing, BIGCOMP 2015	1
International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)	1

International Conference on Computational Intelligence and Communication Networks	1
International Conference on Education, Management, Arts, Economics and Social Science.	1
International Conference on Information and Knowledge Management, Proceedings	1
International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2016)	1
International Conference on Wireless Communications & Signal Processing (WCSP)	1
International Journal Business Process Integration and Management	1
International Journal of Advanced Research in Computer and Communication Engineering	1
International Journal of Big Data Security Intelligence	1
International Journal of Information Security and Privacy	1
International Journal of Latest Trends in Engineering and Technology	1
International World Wide Web Conference	1
IT Professional	1
ITU Kaleidoscope: Trust in the Information Society (K-2015)	1
Journal of Big Data	1
Journal of Business Research	1
Journal of Decision Systems	1
Journal of Information Systems and Technology Management	1
Journal of International Accounting Research	1
Journal of Medical Systems	1
Journal of Policy History	1
Knowledge-Based Systems	1
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	1
Library trends	1
MATEC Web of Conferences	1
Medicina Clínica	1
Molecular Systems Biology	1
Privacy and identity Management for the Future Internet in the age of Globalisation	1
Proceedings of the 2014 Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, CSIBIG	1
Proceedings of the 8Th European Conference on Is Management and Evaluation (Ecime 2014)	1
Proceedings of the Federated Conference on Computer Science and Information Systems; ACSIS	1
Proceedings of the sixteenth international workshop on Data warehousing and OLAP - DOLAP '13	1
Public Health Reports	1
Revista de Ciencias Económicas	1
Revista de los Estudios de Derecho y Ciencia Política	1
Revista Telos: Cuadernos de comunicación e innovación	1
Science and Engineering Ethics	1
Seventh Latin-American Symposium on Dependable Computing (LADC)	1
Statistical Journal of the IAOS	1
System Sciences (HICSS), 2015 48th Hawaii International Conference	1

Teknokultura: Revista de cultura digital y movimientos sociales	1
Telecommunications Policy	1
The Business Lawyer	1
The Participants of the NSF Big Data Security and Privacy Workshop	1
WIT Transactions on Information and Communication Technologies	1
XXIV Congreso sobre Derecho y Salud, Titulado Innovación y desarrollo tecnológico en el ámbito de la salud	1

Fuente: Elaboración propia

Destaca por encima de todas las demás revistas *Computer Law & Security Review*, con cinco artículos en diferentes ediciones de la revista sobre el tema que se trata de estudiar. Se trata de una revista internacional centrada en la práctica del derecho y la tecnología, consolidada como plataforma para publicación de investigaciones de alta calidad.

Aboga por análisis jurídicos en el campo de las Tecnologías de la Información y la seguridad informática. A día de hoy es el principal periódico de investigación en este ámbito en toda Europa, lo cual queda reflejado en el estudio llevado a cabo, puesto que se catapultó a la primera posición de manera relevante.

Los colaboradores que trabajan y publican artículos en esta revista son profesionales dedicados a las principales firmas de abogados internacionales, especializadas en derecho de la Tecnología de la Información, quienes proporcionan comentarios y análisis esenciales basados en la amplia experiencia en la aplicación de la legislación de TI en la realidad práctica.

También destacan en la clasificación entidades productoras como *Institute of Electrical and Electronics Engineers* o *Association for Computing Machinery*, por sus elevadas cifras, pero debemos tener cuidado con estos datos, ya que se trata del recuento de todos los artículos que ofrecen las distintas revistas que publican estas entidades, y cada una de dichas revistas tienen muy diferentes ámbitos de estudio e investigación, en la tabla están plasmados debido a que se hacía necesario recontar dicho número de aportaciones, pero el estudio de estas entidades se hará en el siguiente apartado del trabajo, específicamente dedicado a las entidades productoras de los mismos.

Como podemos observar, además de las entidades comentadas anteriormente y sus respectivas revistas especializadas, no hay una revista que destaque sobre manera de todas las demás en la publicación de artículos sobre la Privacidad en el ámbito aplicativo del Big Data. La mayoría de los artículos se encuentran completamente dispersos en número y forma no solo alrededor del globo, si no a lo largo y ancho de

todas las revistas de artículos sobre el tema que existen en el mundo. Podemos destacar las siguientes revistas por tener entre tres y dos artículos publicados sobre el tema:

Procedia computer science: Fundada en 2009, se trata de un producto electrónico centrado en la publicación de procedimientos y conferencias de alta calidad académica. Permite una difusión rápida para los investigadores de sus artículos, publicándolos y manteniéndolos en línea gratuitamente en ScienceDirect, para que la información y divulgación de dichos artículos esté disponible para todo el mundo.

Se trata de un servicio que vela encarecidamente por la calidad de los artículos que publican, para lo cual llevan a cabo un análisis de todos ellos en aras de que cumplan con una serie de criterios necesarios en base a la calidad, relevancia y popularidad de los temas tratados. La revista se encuentra indexada en el buscador Scopus, anteriormente analizado en esta investigación académica.

Big Data Research: Revista científica cuyo objetivo es promover y comunicar los avances en la investigación de Big Data, proporcionando un lugar de discusión de calidad para los investigadores, profesionales y diseñadores relativos al mismo. A pesar de que todos ellos están altamente especializados, aportan artículos sobre muy variados temas (siempre aplicando las herramientas de Big Data) tales como social web, finanzas, comercio electrónico, salud, medio ambiente, física, astronomía, química, medicina y descubrimiento de fármacos, bibliotecas digitales, seguridad y administración pública.

International Journal of Managing Projects in Business: Se trata de una revista que persigue cubrir los aspectos teóricos y prácticos de la gestión de proyectos, por lo que se encuentra altamente especializado. El fundamento filosófico de la revista es que la gestión de proyectos se centra principalmente en cómo las personas se organizan para ofrecer beneficios y valor a través de proyectos y programas de proyectos.

Se encuentra alojado en *Emerald Insight Publishing*, una editorial online mundial fundada en 1967, que promueve la defensa de nuevas ideas para la investigación en el ámbito de los negocios y la administración- Actualmente han expandido sus influencias a ámbitos de la salud, asistencia social, educación e, incluso, ingeniería. Se trata de una plataforma web que engloba y organiza unas 300 revistas de investigación, más de 2.500 libros y más de 1.500 casos de enseñanza.

Institute of Electrical and Electronics Engineers Cloud Computing: Se trata de una revista especializada en promover la computación en la nube en todas sus vertientes, lo cual conlleva tareas como actuar como líderes en la difusión y educación de información sobre el tema, coordinar actividades colectivas, gestionan conferencias virtuales o incluso promocionan proyectos de investigación en la nube.

Sus publicaciones se centran en computación en la nube y Big Data. También se encargan de crear podcasts, blogs, artículos y una nutrida comunidad de voluntarios e investigadores que discuten todo el conocimiento relativo a esta área de interés.

Institute of Electrical and Electronics Engineers International Congress on Big Data (Big Data Congress): Se trata de un congreso internacional cuyo objetivo principal trataba de convertirse en foro internacional de exploración mediante debate de nuevas e interesantes ideas de negocio que conllevaran cualquier tipo de servicios de valor agregado. La idea de que Big Data es un factor clave para sacar el máximo provecho a los conocimientos empresariales y mayor economía en los servicios. Dentro de este congreso se celebraron a su vez:

- La Conferencia Internacional sobre Servicios Web (IWEA 2014) del IEEE. Conferencia centrada en el modelado, desarrollo, publicación, supervisión, gestión de diversos tipos de entorno en la nube.

- El 2014 IEEE 11a Conferencia Internacional de Servicios de Computación (SCC 2014), conferencia centrada en la gestión del ciclo de vida de servicios, modelado de empresas, consultoría de negocios, optimización y marketing de servicios.

- El 2014 IEEE 3ª Conferencia Internacional de Servicios Móviles (MS 2014), se trata de una conferencia temática relativa a todos los aspectos concernientes a los servicios móviles, incluyendo un énfasis especial en la concienciación de los entornos sociales virtuales.

5.2.4. Número de artículos por universidad o entidad productora

Tabla 2: Relación de artículos por entidad productora

Entidad Productora del Artículo	Número de Artículos
Institute of Electrical and Electronics Engineers	22
Association for Computing Machinery	6
Chinese Academy of Sciences	3
CSIRO Digital Productivity Flagship	2
Nanjing University	2
Norwegian University of Science and Technology	2
Routledge Journals, Taylor & Francis LTD	2
Universidad Nacional de Educación a Distancia - U.N.E.D.	2
University of Texas	2
Aalto University, Finland	1
Abu Dhabi School of Management	1
Aichi Prefectural University	1
Aljabal Algharby University	1

Amsterdam University of Applied Sciences	1
Anil Neerukonda Inst Technol & Sci, Det CSE	1
Association Schools Public Health	1
Australian National University	1
Azad University Quchan	1
Bangalore Chapter (The Institute of Company Secretaries of India)	1
Beijing Academy of Social Sciences	1
Beijing Normal University	1
Birmingham City University	1
Cambridge University Press	1
Carnegie Mellon University	1
Chiang Mai University	1
China Inst Communications	1
Chong Qing Creation Vocational College	1
Chongqing University of Posts and Telecom	1
Christ University	1
College of Business, Abu Dhabi University	1
Comp Sci & Elect Technol Intellectual Society	1
Cornell University	1
CSIRO Computational Informatics, Australia	1
Dalian University of Technology, Dalian, China	1
Delft University of technology	1
Drexel University	1
Ecole nationale superieure d'informatique Laboratoire des Methodes de Conception des Systemes	1
European Union Agency for Network and Information Security	1
Galgotias University	1
Gazi University	1
Georgia Institute of Technology	1
Griffith University Gold Coast Campus	1
Hacettepe University	1
Hamdard University	1
I + D Almirall S. A.	1
Icahn School of Medicine at Mount Sinai	1
International Business Information Management Association	1
Italian Institute for Privacy and Data Valorization	1
Linklaters LLP	1
Massachusetts Institute of Technology	1
Microsoft	1
Moulay university of Saida	1
Murray State University	1
National Institute of Technical Teachers' Training and Research	1

National Institute Technology Rourkela	1
Nexa Center for Internet and Society (Politecnico di Torino)	1
North Dakota State University	1
Open University	1
Princeton University, New York University	1
Purdue University	1
S. V. National Institute of Technology	1
Simon Fraser University, Canada	1
Springer-Verlag Berlin	1
State University of New York at Binghamton	1
SVS College of Engineering, Coimbatore	1
The Chinese University of Hong Kong	1
The Honk Kong Polytechnic University	1
Tsinghua University	1
Universidad de Castilla-La Mancha	1
Universidad de Huelva	1
Universidad de La Coruña	1
Universidade Federal da Bahia	1
Universität Leipzig	1
Universitat Oberta de Catalunya	1
Universiti Teknologi Mara Perak	1
University Hakodate	1
University of Aizu, Japan	1
University of Auckland	1
University of Calabria	1
University of Calgary	1
University of Copenhagen, Denmark	1
University of Illinois	1
University of Insubria	1
University of Ireland Maynooth	1
University of North Carolina at Greensboro	1
University of Northern Iowa	1
University of Oxford	1
University of Pisa	1
University of Southern California	1
University of Sydney	1
University of Technology, Sydney	1
University of the Aegean	1
University of Toulouse	1
University of Waterloo	1
University Polytechn Bucharest	1
Uniwersytet Zielonogórski	1

UNSW Australia	1
Vellore Institute of Technology	1
Vels University	1
Victoria University	1
Wiley-Blackwell	1
Xidian University, China	1

Fuente: Elaboración propia.

En el estudio bibliográfico hemos encontrado una enorme disparidad de universidades y entidades productoras de estudios e investigaciones sobre la Privacidad en Big Data.

En la mayoría de estas instituciones solo se recoge un artículo relacionado con el tema, representando que no existe una doctrina consagrada en ninguna institución dedicada al estudio de la Privacidad de Big Data, lo cual es bastante lógico teniendo en cuenta lo específico del tema, exceptuando una institución a la que se dedicará una pequeña reseña al final de este apartado.

La lista de instituciones que solo tienen una investigación sobre el tema principal de este Trabajo Fin de Máster son la mayoría, como previamente comentábamos, la Universidad de Calabria, en Italia, la Universidad de Ciencias Aplicadas de Ámsterdam, en Holanda, la Academia de las Ciencias Sociales de Beijing, en China, la Universidad de Princeton, en Nueva York, la Universidad de Toulouse, al sur de Francia, la Universidad de Aizu, en Japón.

También cuentan con su granito de arena en el asunto universidades prestigiosas como la Universidad de Oxford, las Universidad del Sur de California o la Universidad de Illinois o la Universidad de Tecnología en Sidney, Australia.

La Universidad de Texas y la Universidad Noruega de la Ciencias y Tecnologías, en Trondheim, destacan sobre las demás por tener dos estudios referentes al tema que nos ocupa cada una, demostrando así un interés mayor por esta preocupación.

En cuanto a Universidades españolas, podemos encontrar la Universidad de Castilla La Mancha, la Universidad de Huelva, la Universidad de La Coruña y la Universitat Oberta de Catalunya, como referentes dentro de nuestro país, con una aportación cada una de ellas a la doctrina internacional.

Destaca principalmente la Universidad Nacional de Educación a Distancia con una aportación de dos artículos académicos por sus docentes productivos.

No solo instituciones dedicadas a la docencia y divulgación del conocimiento se han atrevido a publicar artículos sobre la Privacidad en el manejo de herramientas de Big Data, también se encuentran instituciones privadas que han realizado actividades como la Escuela de Negocios de Abu Dhabi, en los Emiratos Árabes Unidos y la empresa “The Institute of Company Secretaries of India”.

También empresas muy cercanas al mundo universitario, como puede ser la Cambridge University Press, la editorial de la Universidad de Cambridge realiza su actividad junto a la prestigiosa universidad del mismo nombre.

Una empresa privada de nuestro país que ha intervenido en el proceso de estudio de la Privacidad en Big Data es *I + D Almirall S. A.*, establecida en Barcelona, una empresa de investigación que opera en la industria farmacéutica, a través de la mejora y desarrollo de medicamentos.

La empresa Linklaters LLP, despacho de abogados de fama mundial, con sede en diferentes países del mundo, también cuenta con su aportación al tema con su artículo “*Is Big Data creepy?*” (2013).

Mención especial se merecen Microsoft y la Universidad Purdue, que en colaboración consiguieron finalizar el artículo “A Framework for Composition and Enforcement of Privacy-Aware and Context-Driven Authorization Mechanism for Multimedia Big Data” (2015), siendo un fiel reflejo de cómo una colaboración profesional entre una universidad y una empresa multinacional de prestigio centrada en el entretenimiento pueden unir fuerzas para crear doctrina e investigación sobre temas de interés general.

Otra empresa privada que ha llevado a cabo investigaciones sobre el asunto , en concreto 2, uno en 2014 y otro en 2016, se trata de la empresa Taylor & Francis Group, fundada en 1798 y situada en la ciudad de Abingdon, Inglaterra, dedicada a la publicación de revistas y libros académicos.

Springer Science + Business Media o Springer es otra editorial, esta vez alemana, que publica libros, libros electrónicos y publicaciones científicas de revisión por pares relacionados con ciencia, tecnología y medicina. También otra editorial llamada Wiley-Blackwell, situada en Nueva Jersey, Estados Unidos, es conocida por aportar un artículo académico al estudio de la Privacidad en Big Data.

Pero de entre todas estas entidades, existe una a la que se procede a continuación a estudiar debido a su enorme incidencia y aportación en el tema, el resultado de este estudio bibliográfico demuestra que la institución privada “*Institute of Electrical and Electronics Engineers*”, más conocido como por sus siglas IEEE, ha

conllevado a la expansión y desarrollo de conocimientos sobre la Privacidad en el desarrollo y utilización de herramientas Big Data más que ninguna otra empresa, en concreto, entre 22 artículos fueron producidos y/o llevados a cabo bajo el resguardo de esta empresa.

Se trata de una asociación a escala mundial de ingenieros, dedicada a la estandarización y desarrollo de herramientas y licencias en el área técnica de producción. La forman alrededor de 400.000 miembros en más de 150 países y desde hace años se enorgullecen en poseer el título de la mayor asociación internacional sin ánimo de lucro formada por profesionales de las nuevas tecnologías.

Fue creada en el año 1884, y en la creación influyeron personalidades históricamente conocidas como Thomas Alva Edison o Alexander Graham Bell. Desde entonces su actividad principal es dedicada a promover la creatividad, desarrollo e integración de avances tecnológicos en la electrónica y las ciencias en general.

Para ello llevan a cabo actividades de publicación técnica, conferencias de personalidades en las materias específicas y fomentando el consenso entre disciplinas, lo cual les lleva a producir más del 30 % de la literatura especializada sobre ingeniería eléctrica, computación y telecomunicaciones, lo cual nos hace entender el por qué se encuentra en una situación tan destacable dentro de nuestro ranking de publicaciones por entidades.

En menor medida que esta institución, pero también destacando sobre las demás instituciones productivas de artículos sobre el tema de interés, es necesario exponer una breve referencia de las siguientes entidades:

Association for Computing Machinery: En esta institución que se estudia, con 6 artículos desarrollados, se trata de la mayor sociedad de computación del mundo, fundada en 1947. Publica varias revistas y periódicos científicos relacionados con la computación; patrocina conferencias en varias áreas del campo y otros eventos relacionados con las ciencias de la computación y además cuenta en su operatividad práctica con hasta 34 grupos de interés especial.

Chinese Academy of Sciences: Esta institución se trata de una red de investigación y desarrollo integral, una sociedad basada en el mérito y un sistema de educación superior, reúne a científicos e ingenieros de China y del mundo entero para abordar problemas teóricos y aplicados utilizando enfoques científicos y de gestión de clase mundial.

A modo de conclusión, como se puede observar, las instituciones que se han encontrado en este estudio, y que han resultado esclarecedoras y colaboradoras en la materia de estudio, se encuentran dispersas por todo el globo, por lo que se puede decir que es un tema de interés y preocupación de nivel mundial, y a la vez, una materia en la que se avanza y desarrollan ideas de forma unánime y progresiva por todas las mentes pensantes especialistas en el tema; además de ello, podemos destacar que no es una doctrina meramente estudiada por ámbitos puramente universitarios, sino que además ha sido manejado y estudiado por entidades privadas, tanto lucrativas como no, y que cuando se realiza colaboración entre ambas entidades comentadas, se puede llegar a relaciones muy productivas.

5.2.5. Estudio de artículos por autores

Tabla 3: Tabla de autores con dos artículos

Autor	Recuento
Amine, Abdelmalek	2
Bull-berg, Heidi	2
Chen, Jinjun	2
Cuzzocrea, Alfredo	2
Dou, Wanchun	2
Guo, Song	2
Li, Peng	2
Liu, Chang	2
Nepal, Surya	2
Nissenbaum, Helen	2
Olsson, Nils O.E	2
Perera, Charith	2
Rahmani, Amine	2
Ranjan, Rajiv	2
Rousseaux, Francis	2
Saurel, Pierre	2
Wang, Lizhe	2
Wang, Zhong	2
Yang, Chi	2
Yu, Qian	2
Zhang, Xuyun	2

Fuente: Elaboración propia

Exponer en este apartado todos los autores que han participado en cada uno de los artículos de esta revisión bibliográfica conllevaría adjuntar una tabla relación demasiado extensa, pero junto a estas palabras se encuentra adjunta una relación de los autores que han participado más de una vez en algún artículo relacionado con la Privacidad en la aplicación de herramientas productivas de Big Data, sobresaliendo de

algún modo por la media de todos los demás autores de artículos, que solo han aportado una vez a esta doctrina de conocimiento.

5.2.6. Estudio de las Keywords encontradas en la revisión bibliográfica

Para este apartado se analizó y conformó una clasificación de los keywords que aparecían en cada uno de dichos artículos estudiados, contando los mismos para determinar cuáles son más o menos repetitivos y redundantes. Se establecen mediante grupos de tablas donde se han recogido desde los más repetidos a los menos, y más tarde clasificados por grupos de interés, de interés futuro, métodos de privacidad, de ámbito informático, de ámbito legal, ámbito empresarial, ámbito ético o filosófico, ámbito Marketing y de riesgo u oportunidad, eligiéndose los mismos por ser representativos de todos los ámbitos doctrinales encontrados en las palabras claves del estudio y dando valor de la variedad de aplicaciones y ámbitos dónde influye el Big Data.

Procedemos a continuación a estudiar y analizar dos clasificaciones de palabras claves, la primera de ellas mediante la creación de tres escalones de keywords, divididos por cantidad de veces que se encuentran mencionados en los artículos y material académico investigado. Después de dicha clasificación se procede a exponer una serie de tablas con agrupaciones de palabras clave que se pueden asociar a un mismo ámbito aplicativo, otorgando al lector una perspectiva de la cantidad de ámbitos aplicativos en los que se ejerce la privacidad en Big Data.

Tabla 4: Keywords más comunes (o de primer nivel)

Keyword	Recuento
Big Data	71
Privacy	29
Data privacy	23
Cloud computing	10
Anonymization	6
Security	6

Fuente: Elaboración propia

La cúspide de la clasificación la ocupa, como no podía ser de otra forma, el término “*Big Data*”, con 71 menciones a lo largo de todos los artículos, es lógico pues ya que es el tema central sobre el que pivotan los artículos analizados, el hecho de que el número de menciones no sea el de 106 es extraño, pero se puede explicar debido a que no todos los artículos utilizados en este estudio dieron la suficiente información como los keywords que incluían, bien por ser datos reservados o porque directamente no se incluyeron por el autor; otro motivo es que cuatro de los archivos estudiados eran libros, los cuales no suelen ir aparejados de una serie de keywords, y además muchos artículos eran muy específicos, como los dedicados exclusivamente a analizar el uso de

herramientas como MapReduce o Apache Hadoop, en los cuales no encontramos palabras clave como Big Data, pero si palabras clave más específicos de su área de investigación.

En segundo lugar, encontramos los términos “*privacy*” y “*data privacy*”, obteniendo 29 y 23 resultados respectivamente, por ser el término acuñado junto con el tema de opinión Big Data analizado en este estudio, es lógico que también ocupe una posición alta en la clasificación.

En una segunda posición, se encuentra el término “*cloud computing*”, que hace referencia a un tipo de computación distribuida a gran escala sin necesidad de conocimientos o de ser expertos en la gestión de los recursos que se necesitan para ello, que se ha convertido en la fuerza impulsora de las tecnologías de la información y la comunicación en los últimos tiempos (Jain et al., 2016). Muchos de los servicios que los dispositivos electrónicos que utilizábamos décadas atrás requerían de aplicaciones independientes (reproductores de música y video, programas de mensajería instantánea, etc...), pero hoy día todo ello está centralizado en diferentes herramientas, como por ejemplo un navegador web, desde el cual podemos hacer todo ello, a ojos de los usuarios comunes de Internet, este es el reflejo más común del cloud computing (LAMATA, 2012).

Como es lógico es un tema bastante conectado con Big Data, que en la actualidad requiere de herramientas que le posibiliten tanto hacerse efectivo como recolectar datos a gran escala, y la computación en la nube le ofrece la posibilidad de ello de forma flexible y personalizada, por eso encontramos esta palabra clave tan comúnmente en el ámbito de la Privacidad en Big Data.

En el siguiente escalafón se encuentra el término “*anonymization*”, que se trata de una herramienta para mitigar los riesgos que presenta la obtención y tratamiento masivo de los datos de carácter personal, el proceso permite identificar y ocultar la información sensible contenida en los documentos, permitiendo su tratamiento y divulgación sin que ello implique vulnerar la intimidad de las personas u organizaciones a las que hagan referencia los documentos.

Como se puede observar, es una de las herramientas que se encuentran en el desarrollo del contenido de este estudio que permiten soportar y otorgar cierta Privacidad a los datos manejados en Big Data, junto con la ayuda de Hadoop y Mapreduce y su sistema de distribución de datos, más adelante explicado.

También en este mismo nivel se encuentra el término “*security*”, que se encuentre repetida en tantas ocasiones hace referencia a los artículos que pretenden mejorar la seguridad con la que se guardan los datos sensibles con los que se trabaja en Big Data, lo cual es una característica básica si queremos salvaguardar la privacidad de los mismos. Éste y el término previo tienen ambos seis repeticiones dentro de la clasificación de todos los artículos.

Tabla 5: Keywords menos comunes (o de segundo nivel)

Keyword	Recuento
Big Data analytics	5
Cryptography	5
Data Handling	5
Data Mining	5

Fuente: Elaboración propia

En un tercer grupo dentro de la clasificación, se encuentran términos como “*Big Data analytics*”, “*cryptography*”, “*data handling*” y “*data mining*”, todas ellas con cinco repeticiones en todos los artículos estudiados y clasificados. La primera hace referencia al estado de la cuestión de las herramientas analíticas orientadas al tratamiento de datos en Big Data. La segunda se trata de una técnica para preservar la Privacidad de los datos en Big Data, encargándose del cifrado y descifrado del conjunto de datos.

En cuanto a los términos “*Data mining*” y “*Data handling*”, que son procesos sinónimos, se trata del “proceso mediante el cual se trata de descubrir patrones y conocimientos interesantes, útiles y prácticos a partir del desglose y tratamiento de las grandes cantidades de datos” (Xu, Chunxiao, Wang, Yuan, & Ren, 2014). Se podría decir que es el término que describe la actividad más común o la forma de llamar más comúnmente aceptada por la doctrina general al ejercicio de obtener datos interesantes dentro de la marea de datos e información que nos puede aportar el Big Data.

Tabla 6: Keywords de tercer nivel

Keyword	Recuento
Big data applications	4
Big Data security	4
Open Data	4
Privacy Preservation	4

Fuente: Elaboración propia

En un nivel un poco más inferior al comentado anteriormente, se encuentran las palabras clave repetidas cuatro veces a lo largo del estudio de los artículos, que son las siguientes a estudiar: “*Big data applications*” haciendo referencia a las multitudinarias y diferentes aplicaciones en las cuales se puede encontrar Big Data aplicado a nuestro día a día, desde el uso comercial del que hacen algunas empresas para obtener los clientes que más les pueden beneficiar para otorgarles publicidad individualizada en su navegador dependiendo de las búsquedas que haya realizado anteriormente, hasta la utilización por parte de un Estado para indagar en diferentes estadísticas de la población en general.

El siguiente término con dicha cuantía de repetición se trata de “*Big Data security*”, dejando entrever que muchos artículos sobre el tema de la privacidad se preocupan por la seguridad que pueden ofrecer nuestros datos cuando son manejados a escala masiva por las herramientas de Big Data. Del mismo modo, solo que en una vertiente más defensiva académicamente hablando, encontramos el término “*data protection*” que hace referencia a las herramientas que más se utilizan para otorgar de seguridad y protección a los datos en Big Data (P. Li & Guo, 2014).

“*Open Data*” hace referencia a los datos que son o deben ser accesibles de forma gratuita o a un coste mínimo, además permiten acceder a cualquier persona y se pueden utilizar para cualquier propósito sin ningún tipo de restricción de ámbito o persona. Algunos artículos se preguntan los riesgos de hacer públicos datos personales de cantidades enormes de ciudadanos de diferentes países y cómo ello puede influir en la vida de los mismos (Hardy & Maurushat, 2016).

“*Privacy preservation*” es otro de los términos igualmente citados y con dicho número de repeticiones, es lógico que en este ámbito la preocupación por la preservación de la Privacidad esté en auge, por ello muchos artículos tratan de desglosar y estudiar las metodologías existentes para salvaguardar la Privacidad de los datos, como la Anonimización, anteriormente comentada, y de evaluar la eficacia que tienen cuando se implementan y los errores en los que incurren (Shrivastva, Rizvi, & Singh, 2014).

Además de todos estos keywords anteriormente comentados, que son los más comunes y encontrados dentro del estudio bibliográfico, ahora procederemos a estudiar algunos otros que no son tan comunes, pero que han destacado de algún modo dentro de un ámbito aplicativo común.

Tabla 7: Keywords de interés futuro

Keyword	Recuento
Artificial intelligence	2
Authentication in cloud	1
Data warehousing	1
Internet of things	3
Smart cities	1
Palmprint recognition	1

Fuente: Elaboración propia

Términos tales como “*artificial intelligence*”, “*authentication in cloud*”, “*data warehousing*”, “*face recognition*”, “*internet of things*”, “*research institutes*”, “*smart cities*”, “*digital arc*”, “*technological change*”, “*palmprint recognition*” o “*Big Data era*” sirven para tomar conciencia de que Big Data es un denostado ámbito de interés, no solo en los últimos años, sino de cara al futuro próximo, un paradigma aún por descubrir con enorme potencial de desarrollo, muy vinculado a temas de actualidad o que, a su vez, también se encuentran en un estado muy primerizo, tanto de su contexto como de su desarrollo teórico.

Tabla 8: Keywords métodos de Privacidad

Keyword	Recuento
Anonymity	3
Authentication	2
Hadoop	2
Olap	1
Encryption	3
Fragmentation	1

Fuente: Elaboración propia

Otros términos, tales como “*anonymity*”, “*authentication*”, “*criptography*”, “*data complexity*”, “*de-identify*”, “*DNA microarray analysis*”, “*encryption*”, “*feature extraction*”, “*fragmentation*”, “*hadoop*”, “*homomorphic*”, “*olap*”, “*packetization*”, “*distributed databases*”, “*hybrex*”, “*k-anonymity*”, “*mapreduce*” o “*collaborative integration*” hacen referencia todos ellos a estudios, métodos o formas de almacenar, tratar y dar Privacidad a los datos en Big Data de diferentes formas y maneras.

Gracias a ellos se puede otorgar una visión de lo complejo del tema y su aplicación, de las diferentes formas y tipos de métodos existentes para dar solución al problema y de cómo se trata de estudiar y desarrollar cada uno de ellos de cara a su mejora.

Tabla 9: Keywords de ámbito informático

Keyword	Recuento
Internet	3
Computational complexity	2
Personal computing	1
Clustering algorithms	1
Monitoring	2
Cookies	1

Fuente: Elaboración propia

En la tabla número 9 encontramos keywords relacionados con un tema muy vinculado con el ámbito informático e incluso de la programación informática, como dan a entender otras keywords tales como “*convex functions*”, “*anti-tracking*”, “*technologic*”, “*cookies*”, “*clustering algorithms*”, “*technological change*”, “*file servers*”, “*personal computing*”, “*computational complexity*”, “*secure multi-Party computation (SMC)*”, “*electronic devices*”, “*wireless*”, “*monitoring*”, “*machine learning*”, “*java*”, “*clonalg*”, “*nosql datastores*” o “*internet*”. La relación entre todas estas palabras clave y la Privacidad en Big Data es debido a que la programación es la principal herramienta para defender los intereses de los datos personales, por eso muchos de ellos tratan de crear sistemas, algoritmos y mecanismos informáticos que actúen como barrera contra los agentes externos que pretenden violar la intimidad de los ciudadanos y las empresas.

Tabla 10: Keywords de ámbito gubernamental o legal

Keyword	Recuento
Governance	1
Law	2
Privacy rule	1
E-government	2
Privacy policies	1
Rights of privacy	1

Fuente: Elaboración propia

También se han incluido en el análisis keywords que abundan en el ámbito gubernamental de los diferentes estados del mundo, e incluso en el apartado legal del tema, tratando de conseguir una legislación proclive a la defensa de los intereses de la privacidad de los datos. Son un ejemplo de ello “*e-government*”, “*governance*”, “*law*”, “*measures*”, “*privacy invasión*”, “*privacy rule*”, “*research institutes*”, “*national statistical institutions (NSIs)*”, “*decision support systems*”, “*social media*”, “*totalitarian government*”, “*formal verification*”, “*privacy policies*”, “*derecho a la intimidad*”, “*GDPR*” (*general data protection regulation*), llegando a veces a mencionar estados concretos como “*united states*” o “*rwanda*”, así como palabras clave vinculadas a administraciones públicas o políticas de gobiernos estatales tales como “*corporate law*”, “*open government*”, “*privacy regulations*”, “*rights of privacy*” o “*legal aspects*”.

Tabla 11: Keywords de ámbito empresarial

Keyword	Recuento
Companies	1
Market research	1
Managers	1
Business models	2
Commerce	1
Supply chain management	1

Fuente: Elaboración propia

Por supuesto también se encuentran keywords que vinculan el tema con las ciencias económicas y el mundo empresarial de los negocios, tales como “*market research*”, “*stakeholders*”, “*supply chains*”, “*competitive advantage*”, “*accountability*”, “*bankruptcy*”, “*auditor fees*”, “*key management*”, “*auditing*”, “*continuous auditing*”, “*managers*”, “*data marketplaces*”, “*commercial transactions*”, “*logistics*”, “*commerce*”, “*business models*”, “*project management*”, “*supply chain management*”, “*storage management*”, “*decision support systems*” o “*companies*”. Son claros ejemplos de que, además de estar vinculado con el ámbito informático, telemático y de programación, también la Privacidad en Big Data está muy vinculado al mundo empresarial y de los negocios.

Tabla 12: Keywords de ámbito ético o filosófico

Keyword	Recuento
Ethics	1
Tecnologías del yo	1
Etical foresight	1
Bioethics	1
Value of Big Data	1
Ecology	1

Fuente: Elaboración propia

La preocupación por la intimidad y privacidad personal de los datos también lleva a reflexionar internamente a los estudios que ofrecen sus conocimientos para incrementar los estudios sobre de este tema, así encontramos keywords como “*ethics*”, “*personally identifiable information*”, “*sensitivity*”, “*tecnologías del yo*”, “*bioethics*”, “*etical foresight*”, “*information ethics*”, “*medical ethics*”, “*abstracting*”, “*thought control*”, “*ecology*”, “*ecosystem structure*”, “*impacts*”, “*bio-inspired*”, “*social mining*”, “*tecnologías del yo*”, “*healthcare*”, “*control social*” o “*value of Big Data*”, todos ellos evidencian una preocupación ya no solo personal y emotiva, sino también a nivel colectivo como sociedad, del impacto que puede llegar a tener el tema [en ámbitos sanitarios por ejemplo (Mittelstadt & Floridi, 2016)], también se evidencia cierta preocupación por el ecosistema, y un poco de influencia filosófica a la hora de ahondar en la privacidad de los datos (Campbell & Cowan, 2016).

Tabla 13: Keywords de ámbito Marketing y Comunicación

Keyword	Recuento
Market research	1
Questionnaires	2
Freedom of information	1
Conferences	1
Social media	2
Fads	1

Fuente: Elaboración propia

También se encontraron algunos keywords relacionando la Privacidad de Big Data con el mundo del marketing, como los siguientes “*market research*”, “*data pipeline*”, “*sensitive information*”, “*online social network*”, “*information and communication*”, “*sampling*”, “*interview*”, “*questionnaires*”, “*fads*”, “*conferences*”, “*open datum*”, “*freedom of information*”, “*network*”, “*social media*” o “*electronic data interchange*”.

Tabla 14: Keywords de Riesgo u Oportunidad

Keyword	Recuento
Challenges	2
Opportunities	1
Risk management	2
Bankruptcy	1
Computational complexity	2
Evaluation	2

Fuente: Elaboración propia

Por último, algunos keywords destacados encontrados en el estudio pueden ser los relativos al riesgo en el que se incurre con este tema, los peligros que podemos encontrar por lo poco desarrollado o prematuro del concepto, tales como “*challenges*”, “*evaluation*”, “*opportunities*”, “*risks*”, “*abstracting*”, “*public risks*”, “*computational complexity*”, “*risk management*” o, incluso, “*bankruptcy*”.

La primera conclusión que se puede obtener sobre el tema es que en este ámbito los keywords son bastante dispares, y solo unos pocos son capaces de repetirse de manera constante y destacable, de modo que se puede decir que los estudios acaecidos sobre el tema son muy dispares y se aplican en sectores y ámbitos muy diferentes, de ahí que la mayoría de las keywords analizadas se repitan poco o solo una vez.

A modo de conclusión, se puede determinar que las keywords de este estudio bibliográfico representan, ante todo, la variedad y complejidad del tema que se trata de estudiar, la multitud de opciones y aplicaciones en las que se puede encontrar el tema se ven reflejadas en los diferentes ámbitos que representan las keywords comentadas, yendo desde el ámbito informático hasta el empresarial, tocando por el camino cuestiones de marketing e incluso de ética personal y social.

Destacar que la computación en la nube, la anonimización y la seguridad son términos clave en el proceso de creación de todas las herramientas aplicativas de Big Data, por ello encontramos una gran cantidad de artículos que desarrollan ideas y conceptos que pivotan alrededor de estos conceptos.

Las keywords más repetidas son Big Data, Privacy y Data Privacy, por ser las palabras clave sobre las que pivota todo nuestro estudio, encontramos un reflejo de dichas circunstancias en la tabla clasificatoria básica. Tras esto se procede a desarrollar una clasificación de tres niveles en los que se agrupan las palabras clave más repetidas, que son las relativas a las herramientas que se utilizan para hacer efectiva la privacidad en Big Data en la realidad empresarial, es decir, tales como anonimización, criptografía, etc...

Tras esto se procede a la clasificación por “ámbitos”, de las palabras claves encontradas, como podemos observar, en estos keywords que se exponen el número de repeticiones es muy baja en todos los ámbitos, lo cual representa la gran variedad de circunstancias en las que se puede encontrar la aplicación de Big Data, así como la necesidad de preservar la privacidad de los datos que se manejan.

En cuanto a las clasificaciones concretas, podemos concluir que las de “interés futuro” representan el futuro y las posibilidades de desarrollo de los que goza el concepto en todas sus facetas. El de “métodos de privacidad” ayuda a comprender la cantidad de herramientas que tratan de otorgar seguridad a los datos manejados en Big Data.

En un nivel inferior en cuanto a repeticiones, pero igualmente significativos, encontramos palabras clave de ámbito informático, de ámbito legal, empresarial o contable, ámbito filosófico y de riesgo en el desarrollo de las aplicaciones. En todas estas clasificaciones se encuentran pocas repeticiones, pero nos ayudan a adoptar la perspectiva de la gran cantidad de ámbitos en el que se desarrollan herramientas que defiendan y preserven la seguridad en Big Data.

De esta forma ponemos el broche final al análisis bibliométrico desarrollado sobre la Privacidad en Big Data, clasificado en 6 apartados en los que se tratan de examinar las máximas perspectivas posibles de todos los artículos analizados y clasificados.

CAPÍTULO 6. Conclusiones

Como broche final, y tras realizar un estudio teórico del término, comprender y analizar el mismo, analizar el estado de interés del concepto y llevar a cabo la investigación y estudio mediante análisis bibliométrico, se exponen en este último capítulo las principales conclusiones alcanzadas tras cada uno de los mencionados capítulos o fases de estudio.

En la parte final del capítulo se termina exponiendo las limitaciones encontradas a la hora de realizar el estudio, así como nuevas líneas de investigación que se podrían desarrollar sobre el tema.

Procedemos desarrollar las conclusiones alcanzadas por orden de desarrollo de la investigación:

1: Big Data es una herramienta práctica, que permite a las empresas trabajar y tratar masas ingentes de datos y obtener información valiosa para sus intereses, de tal forma que una vez tratados, puedan ayudar a tomar las mejores decisiones posibles en cualquier ámbito empresarial (Mishra & Singh, 2016).

2: Existen múltiples desafíos a la Privacidad de los datos utilizados en Big Data, en todos los ciclos de vida de formación y adaptación de los mismos (X. Li & Shen, 2016).

3: El derecho aplicable en este sector debe actualizarse urgentemente para dar cabida a un nuevo tipo de protección a los derechos de la información íntima de la persona (Pérez, 2015) de tal forma que no pueda existir ninguna circunstancia en la que se pueda vulnerar la intimidad de ningún ciudadano o empresa debido a obsolescencia del derecho aplicable en la materia.

4: Existen diferentes métodos de preservación de la Privacidad en Big Data, pero a pesar de las mejoras llevadas a cabo en los últimos años de cada uno de ellos, ninguno es definitivo, puesto que el ambiente tecnológico en el que se desarrolla esta herramienta sufre una continua evolución.

Tanto por las partes productivas (las empresas tecnológicas crean continuamente productos con mayores posibilidades de procesamiento tecnológico) como por las partes destructivas (cada día existen más métodos para acceder a la información protegida de los datos en Big Data, puesto que se otorgan más y mejores herramientas tecnológicas a los ciudadanos y las empresas, debido a dichos recursos los entes privados con interés en robar y filtrar información de este tipo cada día cuenta también con más recursos para hacerlo).

Por ello las técnicas de privacidad deben mejorar continuamente para ser capaces de defender y proteger los datos en todos los escenarios posibles de violación de la privacidad de los ciudadanos y las empresas, y para ello se necesita una continua investigación en el tema y una legislación aplicable también actualizada al entorno en el que se aplica (Colombo & Ferrari, 2015).

5: Como consecuencia de lo anterior, encontramos la extraña paradoja de que Big Data otorga una herramienta de posibilidades empresariales y productivas casi ilimitadas en la actualidad, y a la vez crea problemas a su propia utilización y protección debido al desafío de la Privacidad (Barocas & Nissenbaum, 2014).

6: Todo proyecto de Big Data (tanto profesional como académico), debe estar profesionalmente diseñado y metodológicamente cuidado para poder crear conocimiento y que éste obtenga una continuidad y capacidad suficientes para llegar a ser eficaz en la finalidad que persigue (López-Borrull & Canals, 2013).

7: Apache Hadoop y MapReduce son las herramientas más utilizadas y más estandarizadas en el sector de aplicación de Big Data (Zakerzadeh et al., 2015), y se encuentran en continua modificación y mejora para dar soporte tanto al potencial que se espera de estas herramientas como para velar por la protección de la privacidad de los datos que manejan en su actividad (Robak et al., 2016).

8: El derecho aplicable en este sector debe actualizarse urgentemente para dar cabida a un nuevo tipo de protección a los derechos de la información íntima de la persona (Barocas & Nissenbaum, 2014).

9: El estudio bibliográfico de la situación actual del tema ayuda a conseguir una perspectiva amplia en cuanto al número de artículos y demás material académico producidos y publicados por países, revistas o entidades productoras, lo cual se procede a desgranar en las sucesivas conclusiones.

10: Se observa cómo, tanto en el término general (Big Data) como en el término específico (Big Data Privacy), hay dos claras tendencias durante los años que lleva en el mundo de la doctrina de investigación (estudiados en el apartado “Estado de la Cuestión” mediante la herramienta *Google Trends*), una primera en la que el término se encuentra en lo que podríamos llamar “estado de gestación”, que en todas las gráficas observadas rondan entre finales de 2012 y comienzos de 2013, justo cuando comienza a existir más doctrina especializada en el tema, y más tarde comienza la segunda tendencia, la cual se podría denominar “estado de explotación”, es decir, el periodo desde el cual, a pesar de que existan pequeños periodos de depresión, el interés sobre el tema se ha mantenido bastante alto y experimentando un progresivo aumento, alguno

artículos refrendan y defienden esta conclusión en sus propios resultados (López-Borrull & Canals, 2013).

11: En cuanto a la cantidad de búsquedas de un término a otro tan dispares (comparando el número de búsquedas de Big Data frente a Big Data Privacy), se puede llegar a la conclusión de que el término Big Data Privacy no es buscado como un tema de ocio o interés popular, de ahí que se encuentren diferencias tan contrastadas en los resultados de búsqueda de ambos términos, seguramente el término más específico quede relegado a estudios académicos o revisiones bibliográficas como las que trata de llevar a cabo este trabajo, y por ende los resultados en navegadores resultan en cifras mucho más humildes.

12: En general se encuentra un Estado de la Cuestión sobre Privacidad en Big Data bastante favorable a la creación literaria, el interés tanto popular como académico sobre el tema sigue en auge (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016), aunque en años pretéritos haya sufrido ciertas recaídas en algunas de las gráficas observadas, pero siempre manteniendo un progreso al alza más o menos progresivo, lo cual denota un interés por parte de las instituciones y académicos que estudian el concepto y sus vicisitudes, lo que en la realidad se plasma como un mayor número de estudios y artículos, incrementándose año tras año (López-Borrull & Canals, 2013; Monreale, Rinzivillo, Pratesi, Giannotti, & Pedreschi, 2014).

13: El resultado del estudio bibliográfico demuestra que la institución privada “Institute of Electrical and Electronics Engineers”, más conocida como por sus siglas IEEE, ha conllevado a la expansión y desarrollo de conocimientos sobre la Privacidad en el desarrollo y utilización de herramientas Big Data más que ninguna otra asociación, en concreto, más de 20 artículos fueron producido y/o llevados a cabo bajo el resguardo de esta asociación que nos ocupa, tanto en diferentes ámbitos como desde variadas perspectivas.

14: Las instituciones encontradas en este estudio, y que han resultado colaboradoras en la materia de estudio, se encuentran dispersas por todo el globo, por lo que se puede llegar a la conclusión de que es un tema que despierta mucho interés para la mayoría de investigadores, y a la vez, una materia en la que se avanza y desarrollan ideas de forma unánime y progresiva por todas las partes especialistas en el tema.

Además de ello, encontramos que es un tema investigado tanto por ámbitos universitarios como por entidades y sociedades privadas (tanto lucrativas como no), existiendo colaboraciones entre ambos muy productivas, dando ejemplo de colaboración para conseguir objetivos comunes y otorgándonos una perspectiva de la variedad de ámbitos donde interesa investigar y mejorar en la privacidad de Big Data. Es decir, la privacidad de Big Data interesa tanto a universidades, como a empresas privadas, como a asociaciones sin ánimo de lucro.

15: Las keywords de este estudio bibliográfico representan, ante todo, la variedad y complejidad del tema que se trata de estudiar, la multitud de opciones y aplicaciones en las que podemos encontrar el tema.

6.1. Epígrafe 1. Limitaciones del estudio:

Existen muchas dificultades e inconvenientes que acaecen en todos los estudios y proyectos que se llevan a cabo. Es complejo y difícil realizar un estudio habiendo realizado un análisis completo de todos los motivos, circunstancias o datos que pueden torcerse o convertirse en una inconveniencia para el desarrollo del conocimiento que se pretende conseguir.

En este estudio en concreto se encontraron varias de las circunstancias que comentamos, comenzando con el ámbito temporal actual de la búsqueda web de artículos, encontrándose dificultad a la hora de si añadir o no los artículos y material académico relativos al año de ejecución de las búsquedas web, el tiempo otorgó la solución, debido a que se realizaron dos líneas de búsquedas, la primera en abril del mismo año, en la que solo se obtuvieron como resultado 3 artículos realizados sobre el tema, más tarde, entre finales de junio y comienzos de julio de 2017, se realizó una búsqueda de los artículos acaecidos sobre el tema, y los mismos se habían multiplicado exponencialmente.

Esto conllevó a una modificación en las gráficas y tablas de datos, y a apartar del estudio el material relativo al año natural 2017, debido a que las cifras que se obtuvieron en un primer momento desvirtuaban los resultados; es decir, si hubo un aumento paulatino en los últimos años desde 2014, de repente exponer que en el año que nos ocupa solo existen 3 artículos sobre el tema no ayudaría a defender la idea de que el interés sigue en aumento actualmente, lo cual hubiese sido contraproducente.

Otra de las limitaciones al estudio se encuentra en el extenso número de artículos manejados, la cantidad de datos a tratar y las herramientas para llevarlos a buen término, que si bien se ha terminado obteniendo recursos para ello, al comienzo presentó enorme dificultad.

Por último, el conocimiento inicial relativo al objeto de estudio, revistiendo enorme interés pero escapando de los conocimientos del alumno, muchos de los artículos estudiados desarrollan ideas o conceptos muy específicos de doctrinas y ámbitos informáticos, matemáticos o computacionales, que se escapan de la doctrina estudiada en la facultad de ciencias económicas y empresariales y la facultad de derecho, instituciones en las que realizó sus enseñanzas académicas la persona que llevó a cabo este estudio, pero estudiando de la manera más eficiente posible la información se han obtenido los resultados que actualmente refleja el mismo.

6.2. Epígrafe 2. Futuras líneas de investigación:

El ámbito aplicativo desarrollado se encuentra en fase embrionaria, de tal forma que hay margen de maniobra para el desarrollo, investigación y aplicación de mejoras sobre el concepto y sus diferentes aristas. En el ámbito de la Privacidad en Big Data, por ser el que ocupa el estudio, resulta interesante realizar estudio bibliométrico sobre las metodologías que los investigadores de los artículos llevados a cabo utilizaron para recolectar y tratar los datos que manejan para sus conocimientos, puesto que muchos de ellos pertenecen a ámbitos completamente ajenos de investigación, y ayudaría a conformar una metodología común que se pudiese aplicar a los estudios sobre Big Data de una forma estandarizada de cara al futuro.

Otros estudios, aunque estos más puramente teóricos o de contenido, conllevarían el desarrollo de nuevos métodos de protección de la Privacidad de los datos masivos, acordes con el potencial de las herramientas informáticas otorgadas hoy día.

Estudios estadísticos sobre la proyección de la privacidad en las personas también serían muy interesantes, tanto pertenecientes al ámbito académico como no, incidiendo en la utilización de los dispositivos, aplicaciones y herramientas tecnológicas que se ponen a nuestra disposición hoy día, por ejemplo estudiando el grado de intensidad con el que exponemos nuestra privacidad en cada uno de ellos, y si conocemos o no la situación en la que nos encontramos.

CAPÍTULO 7. Bibliografía

- Abdellatif, M., Saleh, I., & Blake, M. B. (2015). JPrivacy: A java privacy profiling framework for Big Data applications. In *10th IEEE/EAI International Conference on Collaborative Computing, CollaborateCom 2014* (pp. 501–502).
- Adluru, P., Datla, S. S., & Zhang, X. (2015). Hadoop Eco System for Big Data Security and Privacy. In *2015 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*.
- Agrawal, D., El Abbadi, A., Arora, V., Budak, C., Georgiou, T., Mahmoud, H. A., ... Wang, S. (2015). Mind your Ps and Vs: A perspective on the challenges of big data management and privacy concerns. In *2015 International Conference on Big Data and Smart Computing, BIGCOMP 2015* (pp. 1–6).
- Ahmad, M., Rashid, T., & Mishra, D. K. (2014). Impact of wireless technology on future of big-data industry. In S. R. Mishra D.K. (Ed.), *Proceedings of the 2014 Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, CSIBIG 2014*. Institute of Electrical and Electronics Engineers Inc.
- Albuquerque Junior, A. E. de, & Santos, E. M. dos. (2015). Adoption of Information Security Measures in Public Research Institutes. *Journal of Information Systems and Technology Management*, 12(2), 289–315.
- Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*.
- Aqeel-Ur-Rehman, Khan, I. U., & Rehman, S. ur. (2016). A review on big data security and privacy in healthcare applications. In *Big Data Management* (pp. 71–89). HIET, FEST, Hamdard University, Karachi, Pakistan: Springer International Publishing.
- Barocas, S., & Nissenbaum, H. (2014). Big data's end run around procedural privacy protections. *Communications of the ACM*, 57(11), 31–33.
- B, C. W., Wu, J., & Jiang, S. (2016). Security, Privacy and Anonymity in Computation, Communication and Storage, 10067(November 2016), 288–296.
- Basso, T., Matsunaga, R., Moraes, R., & Antunes, N. (2010). Challenges on Anonymity , Privacy and Big Data. *2016 Seventh Latin-American Symposium on Dependable Computing (LADC)*, 164–171.
- Bolognini, L., & Bistolfi, C. (2016). Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive

95/46/EC to the new EU General Data Protection Regulation. *Computer Law and Security Review*.

- Bull-berg, H., & Olsson, N. O. . (2015). Use of big data in project evaluations Article information. *International Journal of Managing Projects in Business*, 8(3)(3), 491–512.
- Campbell, D. G., & Cowan, S. R. (2016). The Paradox of Privacy: Revisiting a Core Library Value in an Age of Big Data and Linked Data. *Library Trends*, 64(3), 492–511.
- Carpenter, T., & Jones, K. (2015). Big Data as Complementary Audit Evidence. *Journal of International Accounting Research*, 90(4), 1395–1435.
- Cappello, L. (2017). Big Iron and the Small Government: On the History of Data Collection and Privacy in the United States. *Journal of Policy History* , 29(1), 177–196.
- Bolón-Canedo, V., Sánchez-Marroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems*, 86(May), 33–45.
- Chibba, M., & Cavoukian, A. (2015). Privacy, Consumer trust and big data: privacy by design and the 3 c's. In *2015 ITU Kaleidoscope: Trust in the Information Society (K-2015)* (pp. 235–239).
- Colombo, P., & Ferrari, E. (2015). Privacy Aware Access Control for Big Data: A Research Roadmap. *Big data research*, 2(4), 145–154.
- Congreso, X. I., Information, X. I., & Technology, C. (2016). *Turitec 2016*.
- CSA Cloud Security Alliance. (2012). Top Ten Big Data Security and Privacy Challenges. *Cloud Security Alliance*, (November).
- Cumbley, R., & Church, P. (2013). Is Big Data creepy? In *Computer Law and Security Review* (Vol. 29, pp. 601–609). Linklaters LLP, London, United Kingdom.
- Cuzzocrea, A. (2015). Data Warehousing and OLAP over Big Data: A Survey of the State-of-the-art, Open Problems and Future Challenges. *Int. J. Business Process Integration and Management*, 7(4), 372–377.
- Cuzzocrea, A., Bellatreche, L., & Song, I.-Y. (2013). Data warehousing and OLAP over big data. *Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP - DOLAP '13*, (October), 67–70.
- Daries, J. P., Reich, J., Waldo, J. I. M., Young, E. M., Whittinghill, J., Seaton, D. T., ... Chuang, I. (2014). Privacy, Anonymity, and Big Data in the Social Sciences. *Queue*, 12(7), 30:30–30:41.

- Drogkaris, P., & Gritzalis, A. (2015). A Privacy Preserving Framework for Big Data in e-Government Environments. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9264(September), 5.
- Evfimievski, A. (2002). Randomization in privacy preserving data mining. *ACM SIGKDD Explorations Newsletter*, 4(2), 43–48.
- Froment, F. (2017). Análisis bibliométrico de la interacción profesor- alumno a través de las redes sociales bibliometric analysis of the teacher-student, 53–67.
- Fugkeaw, S., & Sato, H. (2015). Privacy-Preserving Access Control Model for Big Data Cloud. In *2015 International computer science and engineering conference (ICSEC)* (pp. 130–135).
- Fan, L., & Jin, H. (2015). A Practical Framework for Privacy-Preserving Data Analytics. *Www'15*, 311–321.
- Gahi, Y., Guennoun, M., & Mouftah, H. T. (2016). Big Data Analytics: Security and Privacy Challenges. In *2016 IEEE Symposium on Computers and Communication (ISCC)* (pp. 952–957).
- Gathegi, J. N. (2014). Clouding Big Data: Information Privacy Considerations. In Z. Gathegi, JN and Tonta, Y and Kurbanoglu, S and AlUmut and Taskin (Ed.), *Challenges of information management beyond the cloud* (Vol. 423, pp. 64–69). Berlin, Germany: Springer-Verlag Berlin.
- Gheid, Z., & Challal, Y. (2015). An Efficient and Privacy-preserving Similarity Evaluation For Big Data Analytics. In R. Raicu, I and Rana, O and Buyya (Ed.), *2015 IEEE/ACM 8th International conference on utility and cloud computing (UCC)* (pp. 281–289).
- Hardy, K., & Maurushat, A. (2016). Opening up government data for Big Data analysis and public benefit. *Computer Law & Security Review*, 33(1), 30–37.
- Heni, H., & Gargouri, F. (2016). Big data security using fragmentation in Mongo data store. In S. K.S. (Ed.), *Proceedings of the 28th International Business Information Management Association Conference - Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth* (pp. 801–812). International Business Information Management Association, IBIMA.
- Hildebrandt, M. (2013). Esclavos de los macrodatos . ¿ O no ?, 17, 7–26.
- Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 25.
- Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32(4), 363–368.

- Jensen, M. (2013). Challenges of Privacy Protection in Big Data Analytics. *IEEE International Congress on Big Data*, 235–238.
- Jiménez, D. L. (2011). Las cookies como instrumento para la monitorización del usuario en la Red: la publicidad personalizada. *Revista de Ciencias Económicas*, 29(2), 16.
- Jutla, D. N., & Bodorik, P. (2015). Pause: A Privacy Architecture for Heterogeneous Big Data Environments. In K. Ho, H and Ooi, BC and Zaki, MJ and Hu, XH and Haas, L and Kumar, V and Rachuri, S and Yu, SP and Hsiao, MHI and Li, J and Luo, F and Pyne, S and Ogan (Ed.), *Proceedings 2015 IEEE International Conference on Big Data* (pp. 1919–1928).
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59–64.
- Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, 31(3), 471–481.
- Kshetri, N. (2014). Big data's impact on privacy, security and consumer welfare. *Telecommunications Policy*, 38(11), 1134–1145.
- Lamarca, R., & Herrera, D. (2015). Más allá de compartir los datos. *Medicina Clínica*, 144(12), 548–549.
- LAMATA, E. (2012). "Cloud Computing": qué es, para qué sirve y cuáles son sus aplicaciones. Recuperado de: <http://www.eleconomista.es/gestion-empresarial/noticias/4087167/07/12/Cloud-Computing-que-es-para-que-sirve-y-cuales-son-sus-aplicaciones-.html>
- Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). *Privacy, big data, and the public good: frameworks for engagement*. *Privacy, big data, and the public good: frameworks for engagement* (Vol. 41 Abingdo, Oxfordshire, England: Routledge Journals, Taylor & Francis Ltd.
- Li, M., Zang, W., Bai, K., Yu, M., & Liu, P. (2013). MyCloud - Supporting user-configured privacy protection in cloud computing. *ACM International Conference Proceeding Series*, 59–68.
- Li, N., Xu, M., Cao, W., & Gao, P. (2015). Researches on data processing and data preventing technologies in the environment of big data in power system. In *Proceedings of the 5th IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies, DRPT 2015* (pp. 2491–2494). Institute of Electrical and Electronics Engineers Inc.
- Li, C., Zhou, P., & Jiang, T. (2015). Differential Privacy and Distributed Online Learning for Wireless Big Data. In *2015 International conference on wireless communications & signal processing (WCSP)*.
- Li, P., & Guo, S. (2014). Load balancing for privacy-preserving access to big data in cloud. *Proceedings - IEEE Infocom*, (December 2016), 524–528.

- Li, P., Guo, S., Miyazaki, T., Xie, M., Hu, J., & Zhuang, W. (2016). Privacy-Preserving Access to Big Data in the Cloud. *IEEE Cloud Computing*, 3(5), 34–42.
- Li, X., & Wang, S. (2014). Big data and its challenges. In *WIT Transactions on Information and Communication Technologies* (Vol. 46 VOLUME, pp. 799–804).
- Li, X., & Zhang, Z. (2014). Exploit the scale of big data for data privacy: an efficient scheme based on distance-preserving artificial noise and secret matrix transform. in *2014 IEEE China summit & international conference on signal and information processing (CHINASIP)* (pp. 500–504).
- Li, X., & Shen, Y. (2016). The Key Technology Research on Privacy Protection Based on Big Data. In Z. B. You, C. T. Wang, S. L. Hu, T. D. Sun, & X. M. Feng (Eds.), *Proceedings of the 2016 International Conference on mechatronics engineering and information technology (ICMEIT)* (Vol. 57, pp. 204–209).
- Lin, C., Song, Z., Song, H., Zhou, Y., Wang, Y., & Wu, G. (2016). Differential Privacy Preserving in Big Data Analytics for Connected Health. *Journal of Medical Systems*, 40(4).
- Liu, W., Uluagac, A. S., & Beyah, R. (2014). MACA: A Privacy-Preserving Multi-factor Cloud Authentication System Utilizing Big Data. In *2014 IEEE Conference on computer communications workshops (INFOCOM WKSHP)* (pp. 518–523).
- Liu, Y. (2015). Privacy Protection Method in the Era of Cloud Computing and Big Data. *MATEC Web of Conferences*, 22, 1041.
- López-Borrull, A., & Canals, A. (2013). La colaboración científica en el marco de nuevas propuestas científicas: Open Science, e-Science y Big Data. *La Colaboración Científica: Una Aproximación Multidisciplinar*, 1, 91–100.
- Lu, R., Zhu, H., Liu, X., Liu, J., & Shao, J. (2014). Toward efficient and privacy-preserving computing in big data era. *IEEE Network*, 28(4), 46–50.
- Mahalakshmi, C. P. N. (n.d.). Privacy-Preserving Cipher text Multi-Sharing Control for Big Data Storage, 7.
- Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *Information Society*, 32(3), 192–199.
- Mantelero, A. (2015). Smart cities , movilidad inteligente y protección de los datos personales, 21, 37–49.
- Mehta, B. B., & Rao, U. P. (2016). Privacy Preserving Unstructured Big Data Analytics: Issues and Challenges. *Procedia Computer Science*, 78, 120–124.

- Minch, R. P. (2015). Location Privacy in the Era of the Internet of Things and Big Data Analytics. *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, 1521–1530.
- Mishra, A. D., & Singh, Y. B. (2016). Big Data Analytics for Security and Privacy Challenges. In Astya, PN and Swaroop, A and Sharma, V and Singh, M (Ed.), *2016 IEEE International Conference on Computing, Communication and Automation (ICCCA)* (pp. 50–53).
- Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics*, 22(2), 303–341.
- Monreale, A., Rinzivillo, S., Pratesi, F., Giannotti, F., & Pedreschi, D. (2014). Privacy-by-design in big data analytics and social mining. *EPJ Data Science*, 3(1).
- Moon, H., Cho, H. S., Jeong, S. H., & Park, J. (2014). Policy Design based on Risk at Big Data Era: Case Study of Privacy Invasion in South Korea. In Kesselman, C and Chen, P and Jain, H (Ed.), *2014 IEEE International Congress on Big Data (Big Data Congress)* (pp. 756–759).
- Muñoz, L. A., Pedro, M., & Bolívar, R. (2013). Análisis bibliométrico sobre la implementación de las tecnologías de la información y la comunicación en las administraciones públicas: aportaciones y oportunidades de investigación. *Revista Innovar*, 27(63), 141–160.
- Nagajothi, S., & N, R. K. (2016). Data Anonymization Technique for Privacy Preservation Using MapReduce Framework, 5(5), 1012–1018.
- Nils O.E. Olsson;Heidi Bull-Berg. (2012). Use of big data in project evaluations, (Scriven 1991), 1–24.
- Njuguna, C., & McSharry, P. (2017). Constructing spatiotemporal poverty indices from big data. *Journal of Business Research*, 70, 318–327.
- O’Keefe, C. M. (2015). Privacy and Confidentiality in Service Science and Big Data Analytics. In M. Camenisch, J and FischerHubner, S and Hansen (Ed.), *Privacy and Identity Management for the Future Internet in the age of Globalisation* (Vol. 457, pp. 54–70). Berlin, Germany: Springer-Verlag Berlin.
- Pavolotsky, J. (2013). Privacy in the Age of Big Data. *The Business Lawyer*, 69(21), 217–225.
- Perera, C., Ranjan, R., & Wang, L. (2015, July). End-to-End Privacy for Open Big Data Markets (Perera_et_al_2015) .pdf. *IEEE Cloud Computing*.
- Pérez, M. M. S. (2015). Big Data O La Acumulación Masiva De Datos Sanitarios : Derechos En Riesgo, 10–12.

- Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. (2014). Big Data Privacy in the Internet of Things Era. *IT Professional*, 17(3), 32–39.
- Power, D. J. (2016). “Big Brother” can watch us. *Journal of Decision Systems*, 25(March), 578–588.
- Privacy, N., & Kantarcioglu, M. (2014). Big Data – Security with Privacy, 1–9.
- Rahmani, A., Amine, A., & Hamou, M. R. (2015). De-Identification of Textual Data using Immune System for Privacy Preserving in Big Data. In *2015 IEEE International Conference On Computational Intelligence And Communication Technology Cict 2015* (pp. 112–116).
- Rahmani, A., Amine, A., & Hamou, R. M. (2016). Combination of Access Control and De-Identification for Privacy Preserving in Big Data. *International Journal Of Information Security And Privacy*, 10(1), 1–27.
- Revista Publicando. (2015). La detección del fraude contable utilizando técnicas de Minería de datos, 2(5), 103–113.
- Robak, S., Franczyk, B., & Robak, M. (2016). Business Process Optimization with Big Data Analytics Under Consideration of Privacy, 8, 1199–1204.
- Rousseaux, F., & Saurel, P. (2016). In light of the legal debate over personal data privacy at a time of globalized Big Data Making Big Data researchers cooperating with lawmakers to find solutions for the future. In ElBaz, D and Bourgeois, J (Ed.), *2016 International IEEE Conferences On Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing And Communications, Cloud And Big Data Computing, Internet Of People, And Smart World Congress (UIC/ATC/SCALCOM/CBDCOM/IOP/SMARTWORLD)* (pp. 398–403).
- Rousseaux, F., & Saurel, P. (2016). The Legal Debate about Personal Data Privacy at a Time of Big Data Mining and Searching Making Big Data Researchers Cooperating with Lawmakers to Find Solutions for the Future. In *2016 FIRST IEEE International Conference On Computer Communication And The Internet (ICCCI 2016)* (pp. 354–357).
- Sáinz, M., Iñigo, M., Raquel, M., & Sanz, M. (2013). El control de la reputación online para prevenir y gestionar una crisis. *Telos Análisis*, 98. Retrieved from
- Salleh, K. A., & Janczewski, L. (2016). Technological, Organizational and Environmental Security and Privacy Issues of Big Data: A Literature Review. *Procedia Computer Science*, 100, 19–28.
- Samuel, A., Sarfraz, M. I., Haseeb, H., Basalamah, S., & Ghafoor, A. (2015). A Framework for Composition and Enforcement of Privacy-Aware and Context-Driven Authorization Mechanism for Multimedia Big Data. *IEEE Transactions On Multimedia*, 17(9, SI), 1484–1494.

- Sarjapur, K., Suma, V., Christa, S., & Rao, J. (2016). Big Data Management System for Personal Privacy Using SW and SDF. In Satapathy, SC and Mandal, JK and Udgata, SK and Bhateja, V (Ed.), *Information Systems Design And Intelligent Applications, Vol 2, India 2016* (Vol. 434, pp. 757–763). Berlin, Germany: Springer-Verlag Berlin.
- Schadt, E. E. (2012). The changing privacy landscape in the era of big data. *Molecular Systems Biology*, 8(612), 1–3.
- Sedayao, J., Bhardwaj, R., & Gorade, N. (2014). Making Big Data, Privacy, and Anonymization work together in the Enterprise: Experiences and Issues. In H. Kesselman, C and Chen, P and Jain (Ed.), *2014 IEEE International Congress On Big Data (Bigdata Congress)* (pp. 600–606).
- Segura Vázquez, A. (2014). El pastor, el doctor y el Big Data. *Teknokultura*, 11(2), 243–257. Recuperado de: <http://teknokultura.net/index.php/tk/pages/view/opr-210>
- Shrivastva, K. M. P., Rizvi, M. a. A., & Singh, S. (2014). Big Data Privacy Based on Differential Privacy a Hope for Big Data. *2014 International Conference on Computational Intelligence and Communication Networks*, (May), 776–781.
- Shukla, S., & Sadashivappa, G. (2014). A distributed randomization framework for privacy preservation in big data. *Proceedings of the 2014 Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, CSIBIG 2014*, (March 2014).
- Soft, B. D. (2013). ¿Qué es Big Data? Recuperado de: <http://www.businesssoftware.net/que-es-big-data/>
- Sommer, P. DI commentary: Big Data and privacy, 15 *Digital Investigation* § (2015). The Boulevard, Langford Lane, Kidlington, Oxford Ox5 1gb, Oxon, England, Birmingham City University: Elsevier Sci LTD.
- Sudarsan, S. D., Jetley, R. P., & Ramaswamy, S. (2015). Security and Privacy of Big Data. *arXiv Preprint arXiv:1601.06206*, 2(1), 121–136.
- Surendra, H., & Mohan, H. S. (2016). Considerations for Privacy Preserved Open Big Data Analytics Platform. In *2016 International Conference On Computation System And Information Technology For Sustainable Solutions (CSITSS)* (pp. 445–449).
- Tascón, M. (2013). Introducción: Big Data. Pasado, presente y futuro. *Telos: Cuadernos de Comunicación E Innovación*, (95), 47–50. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=4423775&orden=417729&info=link%0Ahttps://dialnet.unirioja.es/servlet/extart?codigo=4423775>
- Terence, C., & Ludloff, M. E. (2011). *Privacy and Big Data. The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference* (Vol. 1).

- Thomas, L. D. W., & Leiponen, A. (2016). Big data commercialization. *IEEE Engineering Management Review*, 44(2), 74–90.
- Thorpe, J. H., & Gray, E. A. (2015). Law and the Public's Health. Big Data And Public Health: Navigating Privacy Laws To Maximize Potential. *Public Health Reports*, 130(2), 171.
- Todman, A. (2016). Privacy in the age of big data: recognizing threats, defending your rights and protecting your family. *Archives and Records*, 37(1), 113–115.
- Wang, H., Jiang, X., & Kambourakis, G. (2015). Special issue on Security, Privacy and Trust in network-based Big Data. *Information Sciences*, 318, 48–50.
- Van de Pas, J., & van Bussel, G.-J. (2014). 'Privacy Lost - and Found?' Some Aspects of Regaining Citizens' Privacy by Means of PET in the age of 'big Data'. *Proceedings of the 8Th European Conference on Is Management and Evaluation (Ecime 2014)*, (January 2014), 278–285.
- Vennila, S., & Priyadarshini, J. (2015). Scalable Privacy Preservation in Big Data A Survey. In Vijayakumar, V and Neelanarayanan, V (Ed.), *Big Data, Cloud And Computing Challenges* (Vol. 50, pp. 369–373). Amsterdam, Netherlands: Elsevier Science BV.
- Thuraisingham, B. M. (2015). Big Data Security and Privacy. In *CODASPY* (Vol. 11, pp. 279–280).
- Terzi, D. S., Terzi, R., & Sagiroglu, S. (2015). A survey on security and privacy issues in big data. *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, (December), 202–207.
- Wang, P., Wang, Z., & Chen, T. (2016). Privacy Protection in Online Social Network in the Context of Big Data. In M. Dong, Q and Guo (Ed.), *Proceedings Of The 2015 3rd International Conference On Education, Management, Arts, Economics And Social Science* (Vol. 49, pp. 461–465).
- Wang, Z., & Yu, Q. (2015). Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures. *Computer Law and Security Review*, 31(6), 782–792.
- Wang, Z., & Yu, Q. (2015). Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures. *Computer Law & Security Review*, 31(6), 782–792.
- Xu, L., Chunxiao, J., Wang, J., Yuan, J., & Ren, Y. (2014). Information Security in Big Data: Privacy and Data Mining. *IEEE Access*, 2, 1149–1176.
- Yu, Y.-C., & Tsai, D.-R. (2016). A Privacy Weaving Pipeline for Open Big Data. In R. Kumar, J. Caverlee, & H. Tong (Eds.), *Proceedings Of The 2016*

IEEE/Acm International Conference On Advances In Social Networks Analysis And Mining Asonam 2016 (pp. 997–998).

- Zaforas, M. (2016). Cassandra, la dama de las bases de datos NoSQL. Recuperado de: <https://www.paradigmadigital.com/dev/cassandra-la-dama-de-las-bases-de-datos-nosql/>
- Zakerzadeh, H. ., Aggarwal, C. C. ., & Barker, K. . (2015). Privacy-preserving big data publishing. *ACM International Conference Proceeding Series, 29-NaN-20*, 1–11.
- Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective big data analysis in continuous auditing. *Accounting Horizons, 29*(2), 469–476.
- Zhang, X., Leckie, C., Dou, W., Chen, J., Kotagiri, R., & Salcic, Z. (2016). Scalable local-recoding anonymization using locality sensitive hashing for big data privacy preservation. In *International Conference on Information and Knowledge Management, Proceedings* (Vol. 24–28–Octo, pp. 1793–1802). Association for Computing Machinery.
- Zhang, X., Dou, W., Pei, J., Nepal, S., Yang, C., Liu, C., & Chen, J. (2015). Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud. *IEEE Transactions On Computers, 64*(8), 2293–2307.
- Zhang, X., Liu, C., Nepal, S., Yang, C., Dou, W., & Chen, J. (2013). SaC-FRAPP: a scalable and cost-effective framework for privacy preservation over big data on cloud. *Concurrency And Computation-Practice & Experience, 25*(18, SI), 2561–2576.
-

