

# APROXIMACIÓN AL GRADO DE CONOCIMIENTO Y APLICACIÓN DE *BIG DATA* EN LAS BIBLIOTECAS UNIVERSITARIAS ESPAÑOLAS

*Ana Belén Ríos Hilario\**

Departamento de Biblioteconomía y Documentación. Universidad de Salamanca.

*Alberto Fraile Sastre\*\**

Departamento de Biblioteconomía y Documentación. Universidad de Salamanca.

**Resumen:** Se analiza el grado de conocimiento e implantación de la tecnología *Big Data* y sus características principales en las bibliotecas universitarias españolas inscritas en REBIUN con el objetivo de observar si estas instituciones se encuentran capacitadas para la utilización y aprovechamiento de las ventajas del tratamiento masivo de datos. Los datos son obtenidos mediante un cuestionario cuya respuesta proviene de fuentes internas de las bibliotecas, a partir de los cuales se establecen una serie de conclusiones junto a unas propuestas de mejora y líneas de trabajo futuras que permitan la correcta implantación, uso y aprovechamiento del *Big Data* en la oferta de servicios y funciones de las bibliotecas universitarias españolas.

**Palabras clave:** Big Data; bibliotecas universitarias.

**Title:** APPROACH TO THE DEGREE OF KNOWLEDGE AND APPLICATION OF BIG DATA IN SPANISH UNIVERSITY LIBRARIES.

**Abstract:** It is analyzed the degree of knowledge and implementation of Big Data technology and its main characteristics in the Spanish university libraries registered in REBIUN, with the objective of observing if these institutions are qualified for the use of the advantages of the massive data processing. The data is obtained by means of a questionnaire, whose response comes from internal sources of the libraries, from which a series of conclusions are established together with proposals for improvement and future lines of work that allow the correct implementation, use and exploitation of the Big Data in the offer of services and functions of the Spanish university libraries.

**Keywords:** Big Data; university libraries.

**Copyright:** © 2020 Servicio de Publicaciones de la Universidad de Murcia (Spain). Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

## 1 INTRODUCCIÓN

El concepto de *big data*, o sus equivalentes como grandes volúmenes de datos, datos masivos o macrodatos, hace referencia a aquellos conjuntos de datos que, debido a su cantidad, complejidad y su capacidad de generación y procesamiento dificulta su tratamiento utilizando las herramientas convencionales de procesamiento, fundamentalmente las bases de datos relacionales y estadísticas (De Mauro y otros, 2016; García-Alsina, 2017).

Inherentemente a la definición se encuentran las características básicas que delimitan el término: volumen, variedad y velocidad. El volumen hace referencia a la cantidad de datos generados y almacenados; la variedad se relaciona con la diversidad de los datos que a su vez proceden de fuentes muy diferentes, y la velocidad responde a que las aplicaciones que analizan estos datos requieren que la velocidad de respuesta sea lo suficientemente rápida para lograr obtener la información correcta en el momento preciso (Hernández-Pérez, 2015).

A estas tres “V” se les han ido añadiendo otras muchas relacionadas con la calidad de los datos, tratamiento y gestión o prestaciones de tecnología y software. Estas características son: veracidad, valor, visualización, verificación, variabilidad y viabilidad (García-Alsina, 2017).

Numerosas empresas, tanto públicas como privadas, grandes y pequeñas, así como organizaciones de todo el mundo requieren guardar y administrar adecuadamente todo este “diluvio” de datos que se generan cada día. Por lo tanto, hablamos de un entorno absolutamente relevante para muchos aspectos que afecta a diferentes campos y sectores. Como ejemplos representativos están: el sector privado, las instituciones científicas e incluso los gobiernos que han encontrado en esta tecnología su mejor baza para optimizar su toma de decisiones y sus resultados. Es

\* anarihi@usal.es

\*\* albertofrailesastre@usal.es

Recibido: 24-07-2019; 2ª versión: 17-11-2019; aceptado: 02-12-2019.

precisamente en este último aspecto donde radica el gran reto al que se enfrenta la tecnología de *big data*: almacenar, buscar, compartir, y valorar “datos hasta ahora infrautilizados o inaccesibles”.

El concepto que hoy día tenemos sobre las bibliotecas universitarias se forma a partir de las modificaciones producidas a finales del siglo XX con la aparición de la sociedad del conocimiento, el aumento de la especialización y la demanda de estudios universitarios por los ciudadanos. Actualmente, estas bibliotecas se constituyen mediante “una combinación orgánica de personal, colecciones e instalaciones cuyo propósito es ayudar a sus usuarios en el proceso de transformar la información en conocimiento” (Gómez, 1996, p. 1).

No es muy abundante la literatura científica que versa sobre la aplicación de esta tecnología en el ámbito bibliotecario. Haciendo un breve estado de la cuestión englobaremos, en primer lugar, todos aquellos textos que tratan de definir el concepto y aplicación de *big data* en el área de las Ciencias de la Información y Documentación en general y en el campo de las bibliotecas en particular (Zetterlund, 2016; Golub y Hansson, 2017; Zhan y Widén, 2017b; Jun Li y otros, 2017; Ramos-Simón, 2017). Tras la lectura de estos textos podemos deducir que “los datos no son nada nuevo para las bibliotecas y las Ciencias de la Información (LIS) y *big data* presenta una expansión cuantitativa de un objeto de estudio ya bien conocido” (Golub y Hansson, 2017). Sin embargo, no existe una definición clara sobre *big data* en el ámbito de estudio (Zhan y Widén, 2017b).

En cuanto a las aplicaciones de los datos masivos en estas organizaciones, tampoco existe un acuerdo general. Es así como según Zetterlund (2016) afirma que “las bibliotecas deben transformar sus colecciones en servicios orientados al usuario y para ello pueden aplicar el *big data* en al menos cuatro casos: recursos, *bench-marking* o evaluación comparativa, toma de decisiones y enriquecimiento de servicios. Por su parte, Golub y Hansson (2017) centran el empleo de los datos masivos en tres áreas específicas: la comunicación científica, el intercambio y la curación de datos. De la investigación desarrollada por Jun Li y otros (2017) se desprende que el marco de aplicación de *big data* en las bibliotecas se puede considerar desde cinco dimensiones: recursos humanos, recursos bibliográficos, soporte tecnológico, innovación de servicios y construcción de infraestructura. Finalmente, Zhan y Widén (2017b) restringen las habilidades de los bibliotecarios en relación con los macrodatos a la gestión de conocimiento e incluye también la privacidad de los datos.

El siguiente bloque de artículos se centran en la aplicación real de *big data* en un tipo específico de bibliotecas (REBIUN, 2016; Rani y Rao, 2017; Zhan y Widén, 2017a; Barashdi1 y Al-Karousi (2019) o un área concreta (Alonso y Vázquez, 2016; Harper y Oltmann, 2017).

Relacionado con nuestro objeto de investigación, las bibliotecas universitarias, Rani y Rao (2017) declaran que son cuatro los ámbitos de aplicación de los datos masivos en este tipo de bibliotecas: desarrollo y conservación de colecciones de conjuntos de datos, estadísticas de uso, gestión de datos de investigación y alfabetización de datos en sus programas de formación. Terminan los autores afirmando que “los bibliotecarios académicos tienen un papel claro en el análisis de los datos para ayudar a las instituciones y los interesados en mejorar los servicios y la calidad de la educación”.

Para la comprensión del objeto de análisis es fundamental el artículo elaborado por Al-Barashdi1 y Al-Karousi (2019) quienes realizan una revisión bibliográfica sobre la aplicación de *big data* en las bibliotecas de investigación. El objetivo principal de este estudio es explorar qué técnicas y herramientas se pueden aplicar en las bibliotecas académicas para analizar esta tecnología y, posteriormente, determinar sus beneficios en estas instituciones. Los resultados obtenidos indican que “a pesar de la gran cantidad de investigación realizada sobre este tema, solo unos pocos estudios analizaron la implicación de *big data* en las bibliotecas académicas, incluidas las herramientas y técnicas de análisis”. Realmente, los artículos examinados se centran más en la definición del término que en su aplicación, y son escasos los textos que examinan el papel que deben jugar las bibliotecas universitarias.

Desde el punto de vista metodológico, es muy importante para este artículo el estudio elaborado por REBIUN realizado en el año 2016 titulado *Big data en las universidades y bibliotecas universitarias españolas*. Aunque íntimamente relacionado con el análisis aquí presentado diremos que hay una diferencia fundamental: el estudio de REBIUN se centra en el conocimiento teórico de las instituciones universitarias en lo referente a los datos masivos, mientras que nuestra investigación pretende dar un paso más y analizar las prácticas reales que sobre *big data* se están llevando a cabo en este tipo de bibliotecas.

Centrado en un caso particular está el artículo titulado *Big Data: la próxima «gran cosa» en la gestión de la información* (Alonso y Vázquez, 2016). Sus autores abordan la alfabetización de datos frente a la alfabetización de la

información tradicional y examinan la primera en el mundo académico, incluyendo una breve reseña bibliográfica de las prácticas de instrucción recientes en este sentido.

En cuanto a la factibilidad referente a la aplicación de *big data* y bibliotecas Ferrer-Sapena y Sánchez-Pérez (2013) afirman que las “bibliotecas de investigación juegan un papel vital en la gestión y curación de este tipo de contenido, pero requieren de mecanismos de financiación adecuados”. En este mismo sentido, podemos decir que “la mayoría de las bibliotecas piensan que el marco de aplicación de *big data* es factible”. Por el contrario, se citan como principales obstáculos de empleo los recursos humanos y el nivel de tecnología de la información (Jun Li *et al.*, 2017). Otro aspecto interesante dentro de los macrodatos es el de su sostenibilidad (Erway y Rinehart, 2016). En este sentido los autores afirman que “debido a que algunos datos de investigación son un valioso activo para la universidad, las instituciones deben construir mecanismos de financiación continua reflejados en sus presupuestos base para proporcionar recursos a las unidades responsables de la gestión de ese activo”.

No nos centraremos, porque quedan fuera de nuestro alcance de la investigación, en aquellos documentos centrados en el nuevo perfil del bibliotecario acorde a esta nueva tecnología, sirvan de ejemplo el artículo de Serrano-Cobos (2013) y el de Harper y Oltmann (2017), ni en la descripción de aplicaciones concretas fuera del contexto de las bibliotecas universitarias, tal es el caso de la Biblioteca Nacional de Singapur o la Biblioteca Nacional de España.

La elección de las bibliotecas universitarias no ha sido fortuita ya que se ha considerado que son estas quienes incorporan de manera rápida los últimos avances en el sector documental y tecnológico. Además, dichos centros están ligados a las instituciones educativas de ámbito superior que combinan la docencia e investigación y, por tanto, la transferencia del conocimiento. Es así como, nuestro análisis se focaliza tanto en el conocimiento actual del concepto por parte del personal bibliotecario como en el nivel de desarrollo de esta tecnología y en dichas instituciones.

Partiendo de las anteriores premisas, el objetivo principal de este artículo es determinar cuál es el grado de conocimiento de la tecnología *big data* en las bibliotecas universitarias españolas, así como prever su impacto en el futuro cercano y a largo plazo, pudiendo establecerse directrices o propuestas de mejora que fomenten la utilización de este tipo de tecnología. A partir de este objetivo general, se pretende dar respuesta también a otros de carácter más específicos:

1. Analizar la literatura científica que verse sobre el concepto de *big data* aplicado a las instituciones bibliotecarias.
2. Conocer el número de bibliotecas universitarias que están aplicando esta tecnología.
3. Conocer la posición del personal bibliotecario frente a la implantación de los datos masivos.
4. Analizar las formas en las que *big data* se está aplicando en dichos organismos.
5. Realizar recomendaciones y propuestas de trabajo en base a los resultados obtenidos previamente.
6. Establecer las líneas de trabajo futuras en el campo del *big data* en bibliotecas, concretamente en el caso de las de educación superior.

Para dar respuesta a dichos objetivos, se estructura este artículo en cuatro grandes apartados. En primer lugar, el capítulo introductorio en el que se documenta la situación actual del objeto de estudio a través de la definición del concepto en el caso de las bibliotecas universitarias; el establecimiento de los objetivos generales y específicos y detalle de su estructura. En la segunda parte, se establece la metodología llevada a cabo para cumplir con los objetivos propuestos. Se ha aplicado una metodología fundamentalmente cuantitativa a través de aplicación de un cuestionario de preguntas abiertas y cerradas. Dicho apartado se divide en cuatro puntos clave: fuentes documentales, localización de la población objeto de estudio, diseño y envío del cuestionario y recopilación de los datos. Posteriormente, se pasa a analizar los resultados obtenidos, terminando el artículo con la redacción de las conclusiones e indicación de las propuestas de mejora y futuras líneas de investigación.

## 2 METODOLOGÍA

Para la obtención de la bibliografía específica en la materia – *big data* y bibliotecas universitarias – se procedió a la consulta de bases de datos generales y especializadas como *Google Scholar*, *Academic Search Complete*, *Dialnet*, *CSIC*, *Emerald Insight*<sup>7</sup>, *Library & Information Sciences Abstracts*, *Library Information Science & Technology Abstracts*, y *Web of Science*<sup>10</sup>. Los términos empleados para la realización de la estrategia de búsqueda fueron principalmente “*big data*” y “bibliotecas universitarias” y sus equivalentes en otras lenguas. De toda la documentación recopilada, se han citado exclusivamente las obras que se han empleado en el presente artículo.

Teniendo en cuenta que el objeto de aplicación del *big data* es en las bibliotecas universitarias, se procedió acto seguido a su localización. En este sentido la búsqueda se realizó en la web oficial de la Red de Bibliotecas

Universitarias Españolas (REBIUN). En total se obtuvieron 74 bibliotecas universitarias que conformarían la población objeto de estudio.

A continuación, tras decidimos por la encuesta como método de investigación ya que lo considerábamos el más apropiado para la recopilación de la información por parte de la población objeto de estudio, se procedió al diseño y elaboración del cuestionario como instrumento de obtención de datos.

Se utilizó para la creación y envío del cuestionario la plataforma Gmail de Google. Desde dicha aplicación que genera el formulario de consulta se envió por primera vez el día 4 de marzo y por segunda vez el 18 del mismo mes, estableciéndose como fecha límite el día 31.

En lo que respecta al cuestionario, adjunto a este trabajo, el mismo se compone de seis apartados diferenciados en función del contenido de las preguntas o la información que otorgan al lector de este, los cuales son los que aparecen a continuación:

1. Presentación, en la que se pretende introducir a la persona que responde en la investigación, por lo que se explica quién es el investigador, qué proyecto se está realizando, sobre qué temática trata y para qué se utilizaran los datos recopilados.
2. Datos de interés, donde se recogen tres preguntas obligatorias que permiten la identificación de la biblioteca que responde y las circunstancias que la rodean como su área de conocimiento o su relación con la institución académica.
3. Concepto de *big data*, en la que aparecen un total de siete preguntas obligatorias que recopilan información sobre el concepto estudiado que existe en cada institución como, entre otras, si conocen el término, cuáles son sus principales características o si lo consideran beneficioso.
4. *Big data* en la institución, con siete preguntas que permiten conocer el grado de aplicación de las tecnologías de análisis de datos masivos en la biblioteca universitaria o científica que da respuesta al cuestionario.
5. Perspectivas de futuro, con tan solo dos preguntas abiertas y no obligatorias que buscan la visión de las bibliotecas sobre el futuro del término y de la tecnología *big data* en nuestro ámbito profesional, con el objetivo de poder establecer vías de estudio futuras y propuestas de mejora.
6. Una despedida, donde se agradece a la persona que ha respondido su colaboración, esfuerzo y dedicación al responder a las preguntas.

Por último, y tras tener todas las respuestas preparadas para su cómputo, se procedió a la realización de un análisis global de las repuestas y a la elaboración de tablas y gráficos para lo cual se empleó la herramienta Microsoft Excel 2013.

En total se obtuvo respuesta por parte de 18 bibliotecas (24.3%). El resto de la población que compone la muestra se divide de modo equitativo entre los que no cumplieron el cuestionario ya que indicaron su desconocimiento en la materia a tratar y los que omitieron su respuesta directamente. Por lo tanto, podemos considerar la muestra representativa para los objetivos del estudio anteriormente mencionados.

### 3 RESULTADOS Y DISCUSIÓN

Las 18 bibliotecas que participaron respondían al perfil de biblioteca universitaria central, donde destacaba la presencia mayoritaria de aquellas cuyo ámbito científico era multidisciplinar (66.6%), seguidas por las pertenecientes a Ciencias Sociales (22.2%) y, finalmente, las correspondientes tanto a Ingeniería y Arquitectura como a Humanidades (5.5%). En este sentido, aunque cuantitativamente la muestra puede parecer pequeña, en función de los resultados obtenidos podemos deducir que las bibliotecas que han contestado son las que conocen y/o aplican esta tecnología, por lo tanto, los resultados pueden considerarse como ilustrativos y significativos. De este modo, la ausencia de respuesta puede deberse más al desconocimiento de la tecnología que al desinterés en el cumplimiento del cuestionario.

Analizando los resultados obtenidos, las bibliotecas encuestadas demuestran el conocimiento del término *big data* y son capaces de elaborar una definición precisa, lo que refleja la importancia que ha cobrado el término y sus tecnologías en los últimos tiempos. De las declaraciones elaboradas por las distintas instituciones (Tabla I), se pueden extraer una serie de características comunes que se asemejan en gran medida a la definición propuesta al inicio de este trabajo. Dichas propiedades se resumen en: grandes volúmenes de datos; no pueden ser gestionados por las bases de datos tradicionales; son datos carentes de estructura y generados por la actividad humana.

Definición del concepto	Biblioteca universitaria
Análisis y explotación de grandes volúmenes de datos	Universitat Politècnica de València (UPV)
Cantidades masivas de información real que puede ser procesada y analizada	Universitat de València (UV)
Volúmenes ingentes de datos	Mondragon Unibertsitatea (MU)
Conjuntos de datos que por su tamaño, velocidad y complejidad deben ser tratados con ayuda de herramientas informáticas	Universidad de Cantabria (UC)
Son grandes cantidades de información que fluctúan por internet, cuyos datos son recopilados por las empresas para detectar nuevas oportunidades de negocio	Universidad a Distancia de Madrid (UDIMA)
Datos masivos	Universitat Jaume I (UJI)
Grandes volúmenes de datos, generalmente sin estructurar, que contienen información útil	Universidad de Castilla-La Mancha (UCLM)
Macrodatos	Universidad de Alicante (UA)
Gestión de datos masivos, inteligencia de datos y datos a gran escala	Universitat Oberta de Catalunya (UOC)
Datos masivos	Universitat Abat Oliba CEU (UAOCEU)
Almacenar grandes cantidades de datos para analizar los usos y el comportamiento de los clientes	Universidad Complutense de Madrid (UCM)
El entramado de información que subyace en los datos resultantes de nuestras transacciones y movimientos en la red	Universidad de Extremadura (UEX)
Gran volumen de datos de todos los ámbitos de la vida	Universidad de Alcalá (UAH)
Gestión de datos	Universidad de Cádiz (UCA)
Este término define los datos masivos que se producen actualmente y que no pueden ser gestionados mediante las bases de datos tradicionales	Universidad de Las Palmas de Gran Canaria (ULPGC)
La recopilación, análisis y aplicación a distintas finalidades de las grandes masas de datos que genera la actividad humana y que quedan recogidos en soporte informático	Universidad de La Laguna (ULL)
Gestión de grandes volúmenes de información (datos) que crecen rápidamente	Universitat Politècnica de Catalunya (UPC)
Tratamiento automatizado de datos de una determinada actividad, organización... y extracción de la información para la toma de decisiones a partir del producto de ese tratamiento	Universidad Pontificia de Salamanca (UPSA)

Tabla I. Definiciones de *big data*.

Por otra parte, preguntadas por las características más importantes de la tecnología *big data*, y permitiendo elegir más de una característica como respuesta, se registraron 63 votos que se centraron, principalmente, en el volumen (23.8%), el valor (15.9%) y, por último, en la velocidad (14.3%) y la variedad (14.3%). Sin embargo, otras características importantes como la verificación (4.8%) o la veracidad (11.1%) que deben presentar los datos para poder ser utilizados eficazmente han sido relegadas a un segundo plano (Figura 1), a pesar de ser dos de los aspectos donde más pueden contribuir los especialistas en documentación.

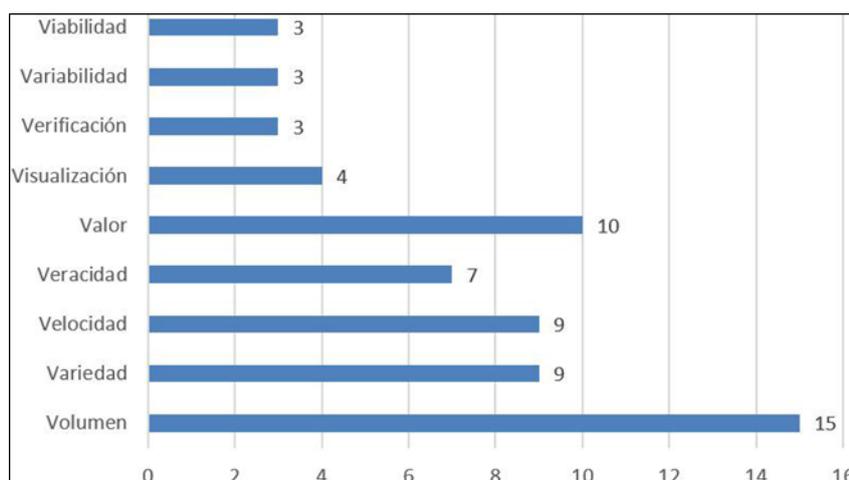


Figura 1. Características más importantes del análisis de grandes volúmenes de datos.

Además, estas bibliotecas estimaron que las fuentes de datos sobre las que las instituciones deben realizar principalmente la recogida, tratamiento y explotación deberían ser las fuentes internas (77.8%). Entienden que los datos generados por la propia institución facilitan el aprendizaje sobre el funcionamiento de la misma repercutiendo, a su vez, en la toma de decisiones para mejorar su atención a usuarios, recursos, etc.

Las fuentes externas (11.1%) son menos valoradas a pesar de que pueden aportar información sobre políticas y servicios que se llevan a cabo en otras bibliotecas y la forma de adaptarlos a la institución. En contraposición, algunas bibliotecas (11.1%) albergaban dudas sobre qué fuente de datos es la más adecuada para el análisis de datos masivos (Figura 2).

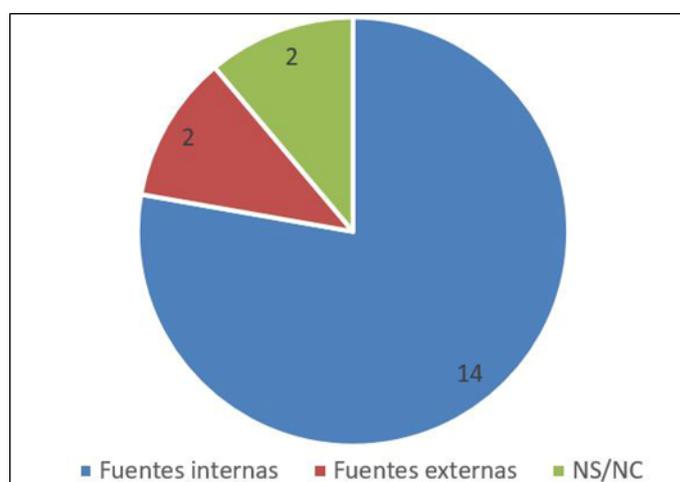


Figura 2. Principales fuentes de datos de *big data*.

De igual manera, se encuentra una clara desigualdad en torno al peso que otorgan a las diferentes fases de la cadena de valor del análisis de *big data*, destacando en primer plano el examen y situándolo muy por encima de la adquisición y almacenamiento y de la generación de datos (Figura 3). Por tanto, se observa como prestan una mayor importancia a la etapa final de la cadena de valor del *big data*, donde finalmente se extrae el conocimiento necesario para la toma de decisiones y la mejora de la institución, siendo este el objetivo de todo el proceso.

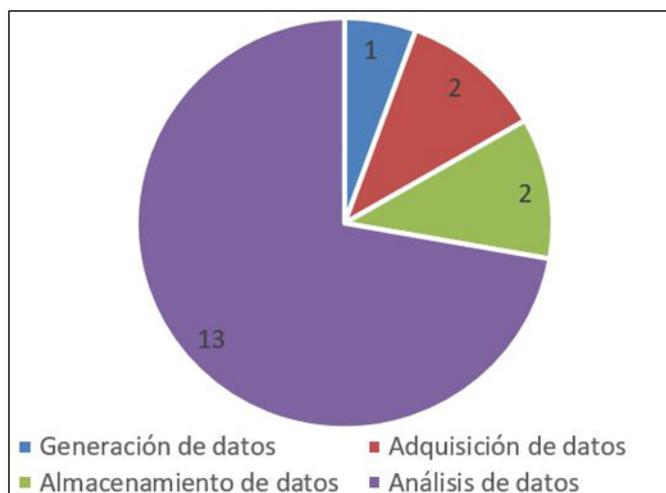


Figura 3. Principales fases de la cadena de valor del análisis de datos masivos.

En general, han considerado que en la actualidad no existe ningún perfil especializado en el análisis de datos masivos (77.7%), hecho que, teniendo en cuenta la relativa novedad del término y especialmente de su aplicación, no es de extrañar que no cuente con un perfil formativo, científico y profesional suficientemente desarrollado. No obstante, mediante el empleo de 51 valoraciones ponen de manifiesto que los profesionales que deben adaptar su ámbito de estudio y trabajo hacia este nuevo perfil deben ser los bibliotecarios (27.5%), estadísticos e informáticos (25.5%) y, en menor grado, los matemáticos (19.6%) y los expertos en Ciencias Sociales (2%) (Figura 4).

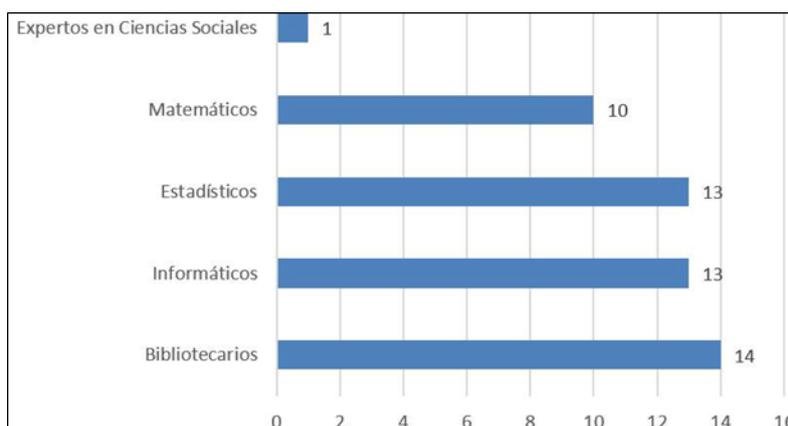


Figura 4. Perfiles susceptibles de convertirse en expertos del big data.

Para ello, proponen la adquisición de una serie de competencias (Tabla II) entre las que destacan por su interés y frecuente repetición:

- Disponer de una gran ética de trabajo para la utilización y explotación de datos sensibles acerca de las actividades y acciones reales que realizan las personas en un determinado entorno.
- Poseer conocimientos de informática y tener dominio del software de tratamiento de datos o conocimientos de programación en diversos lenguajes, con el objetivo de generar la infraestructura digital necesaria para el análisis de datos masivos y su aplicación práctica.
- Disponer de competencias en estadística y tratamiento avanzado de datos, sin los cuales sería inviable la automatización de los cálculos y procesos de manera que puedan ser realizados rápidamente por programas, algoritmos o máquinas.
- Tener habilidades de documentación, búsqueda y filtrado de información, de manera que se asegure el cumplimiento de importantes características del big data como la verificación o la veracidad durante todo el ciclo de vida de los datos.

Competencias de los nuevos perfiles especializados	Biblioteca universitaria
Capacidad analítico sintética, matemáticas, estadística, bases de datos y estructura de la información	UPV
Conocimientos de estadística, informática y documentación	UV
Pluridisciplinar	MU
Informática, matemáticas, éticas y sociológicas	UC
Conocimientos financieros, identificar las necesidades del mercado, saber analizar datos estadísticos, manejo de software especializado en grandes volúmenes de datos, tomar decisiones estratégicas	UDIMA
Saber gestionar los datos desde el inicio hasta su reutilización	UJI
Gestión de gobierno de datos	UCLM
Analistas y técnicos	UA
Dominio del software de tratamiento de datos, dominio de los esquemas de metadatos estándares, pensamiento crítico y resolución de problemas, visualización de datos, capacidad analítica de datos a nivel cualitativo y cuantitativo	UOC
Capacidad de organización, adaptación al cambio y comunicación	UAOCEU
Expertos en biblioteconomía y manejo de bases de datos interrelacionadas	UCM
Búsqueda y filtrado de información, uso ético, creación de algoritmos y tratamiento avanzado de datos	UEX
Ser organizado y tener conocimientos de informática, nuevas tecnologías, formatos y archivos de datos y metadatos	UAH
Conocimiento de mecanismos de actividad investigadora	UCA
Capacidad de análisis estadístico y conocimiento de gestión de bases de datos	ULPGC
Conocimientos de informática, estadística y gestión de la información	ULL
<i>Data mining</i> , bases de datos NoSQL, estadística y gestión de datos	UPC
Conocimientos de programación, mentalidad proactiva abierta y sobre todo un profundo conocimiento de su institución y el tipo de información, usuarios y recursos con los que trabaja	UPSA

**Tabla II. Competencias de los nuevos perfiles especializados en análisis de macrodatos.**

En lo que respecta a estas bibliotecas, la mayor parte (83.3%) opinan que el análisis de los datos masivos, así como sus aplicaciones y tecnologías derivadas presentan numerosos beneficios para sus instituciones frente a aquellas que creen que el *big data* no es útil en su campo de trabajo (11.1%) o que todavía no tienen clara su postura (5.6%).

Entre las principales propuestas de mejora en bibliotecas universitarias mediante el uso de *big data* (Tabla III) destacan por su interés, innovación y posibilidades prácticas las referentes a tres campos principalmente: la atención a usuarios, la optimización de la institución y el apoyo a la investigación.

En el primer caso, el descubrimiento de tendencias en la búsqueda y uso de recursos, el conocimiento de las necesidades reales de los usuarios y la adaptación de los servicios permitirían aumentar el grado de atención personalizada a los usuarios, incrementando a su vez el rendimiento académico o investigador de los mismos y su satisfacción con la actividad realizada por la biblioteca.

Del mismo modo, la optimización de la institución vendría dada por la capacidad de adecuar su plan estratégico a los objetivos marcados por los usuarios y de asignar presupuestos atendiendo a las necesidades detectadas previamente. En cuanto al apoyo a la investigación, el análisis de datos masivos permitiría el reaprovechamiento y preservación de datos al gestionar repositorios donde los autores pudieran incluir tanto las investigaciones como los datos de las mismas, de modo que pudieran ser reutilizados y comparados posteriormente.

Posibles mejoras mediante la utilización de <i>Big Data</i>	Biblioteca universitaria
Descubrimiento de tendencias y comportamientos de usuarios, usos de espacios, servicios y recursos bibliográficos	UPV
Conocer las expectativas y necesidades de los usuarios	UV
Conocimiento e innovación, así como mejores servicios	UC
Conoceríamos mejor las necesidades reales de los estudiantes. Es decir, qué estudios busca, el tipo de aprendizaje que necesita y cómo busca la información	UDIMA
En la gestión diaria y en adecuar más los servicios de la Universidad a las necesidades de los usuarios	UJI
Reaprovechamiento de la información, tanto para la gestión como para la investigación	UCLM
En optimizar el presupuesto de recursos electrónicos y en aumentar la satisfacción de usuarios	UA
Mejorar el apoyo que se ofrece a la investigación, mejorar el cuadro de indicadores de la actividad de la biblioteca dentro de su sector, contribuir a la mejora de la interpretación de los indicadores globales de la institución	UOC
Gestionar correctamente los datos provenientes de la investigación	UAOCEU
Todos nuestros usuarios tendrían acceso a información más pertinente a sus necesidades	UCM
Dar acceso a información relevante para la investigación	UAH
Optimización de la investigación y mejora en la toma de decisiones	UCA
Para la preservación de datos y el análisis estadístico	ULPGC
Adaptación de los servicios, recursos e infraestructuras a las necesidades reales de los usuarios	ULL
En el avance científico, conocimiento y toma de decisiones	UPC
Plan estratégico	UPSA

Tabla III. Posibles mejoras en las instituciones mediante la utilización de *big data*.

En cambio, pese a considerar mayoritariamente que el análisis de datos masivos puede proveer de valor a sus instituciones, muy pocas de estas bibliotecas (16.7%) han adquirido y utilizan algún tipo de tecnología *big data*, encontrando que un gran número de ellas (83.3%) no las han implementado por diversos motivos (Tabla IV). Entre estas causas destacan: la priorización de otras áreas de la biblioteca por parte de la institución universitaria a la que prestan servicio y la falta de apoyo tecnológico para el desarrollo de este tipo de proyectos.

Motivos de no implementar <i>Big Data</i>	Biblioteca universitaria
Por falta de tiempo	UJI
Se está trabajando en ello	UCLM
Gestión interna de prioridades	UOC
Por el tamaño de la institución y el ámbito de investigación	UAOCEU
No se considera prioritario, ni forma parte de la política de las universidades	UCM
No lo hemos incluido entre nuestras prioridades	UEX
No se ha planteado la necesidad	UAH
No tengo respuesta	UCA
No se ha planteado como un objetivo de la biblioteca	ULPGC
Falta de conocimientos especializados y de soporte informático	ULL
Lo estamos intentando pero nos falta apoyo tecnológico y de las propias instituciones, que no ven todavía claramente los beneficios de estos procedimientos	UPSA

Tabla IV. Motivos por los que no se han implementado tecnologías *big data*.

A pesar de esta circunstancia, las reducidas bibliotecas que sí han logrado establecer el análisis de datos masivos entre sus tareas utilizan los siguientes programas (Tabla V):

- *Oracle Business Intelligence Enterprise*: conjunto de herramientas de inteligencia empresarial que permite realizar análisis de negocios.

- *DSpace*: software que facilita la administración de colecciones digitales que se utiliza principalmente para gestionar repositorios institucionales.
- *Apache Solr*: plataforma de búsqueda empresarial de código abierto basada en *Apache Lucene* que posibilita búsquedas facetadas, manejo de documentos enriquecidos y presenta características NoSQL.

Además, desconocen o remiten a otros departamentos los problemas a los que tuvieron que hacer frente durante la puesta en marcha de esta tecnología. Generalmente, fueron los departamentos informáticos de la institución los encargados de la parte técnica del proyecto.

Tecnología de análisis de datos masivos utilizada	Dificultades de implementación	de Biblioteca universitaria
Hacemos explotación estadística de datos a través de herramientas tipo Oracle Business Intelligence Enterprise	Lo desconozco	UPV
Lo está trabajando el servicio de informática	Lo está trabajando el servicio de informática	UV
Solr y el repositorio	Volumen y almacenamiento	UPC

**Tabla V. Tecnologías de análisis de datos masivos que utilizan y dificultades de implementación.**

Finalmente, las bibliotecas determinaron las líneas de trabajo futuro que el *big data* debe establecer en su ámbito laboral (Tabla VI) y que se resumen en los siguientes puntos:

- Asentar el crecimiento conjunto, la colaboración y la interrelación de ambas disciplinas.
- Transformación de la figura del bibliotecario en un especialista en la gestión, preservación y reutilización de datos mediante la creación de nuevos perfiles académicos y profesionales.
- Descubrimiento de tendencias y comportamientos de los usuarios en las búsquedas bibliográficas, uso de recursos, espacios y servicios a través del análisis de datos reales y continuos, permitiendo invertir el presupuesto en los elementos más utilizados y requeridos por nuestros beneficiarios.
- Obtener la capacidad de acceder, procesar, buscar y recuperar a través de una única interfaz en todos los sistemas informáticos de la institución (catálogos, repositorios, bases de datos, etc.).

Líneas de trabajo futuro	Biblioteca universitaria
Descubrimiento de tendencias y comportamientos de usuarios, usos de espacios, servicios y recursos bibliográficos	UPV
Con el <i>Big Data</i> se abre una nueva era en el tratamiento de la información	UV
Colaboración	MU
Avanzar hasta tal punto que las bibliotecas obtengan información real y continua de los perfiles sociológicos de los usuarios y del medio (físico o electrónico) de donde obtienen la mayoría de la información, información que se necesita de una biblioteca para detectar nuevos servicios que se puedan implementar	UDIMA
En la visualización y reutilización de datos	UJI
Las bibliotecas y archivos, así como los profesionales que trabajan en ellas, poseemos el perfil y la capacidad para estructurar los datos y convertir la información en conocimiento	UCLM
Hacia la mejora de la formación de los profesionales y el dominio o adquisición de una herramienta de gestión de <i>Big Data</i>	UOC
Crear la figura de un especialista en su gestión	UAOCEU
Es el futuro de las bibliotecas, el almacenamiento de información en documentos es el pasado, ya que ahora la información está en todas partes	UCM
Interoperabilidad de sistemas informáticos manejados por la institución	UEX
Analizar bien lo que nos interesa y lo que no	UAH
Nueva área de trabajo a implementar	UCA
La relación tendría que ser cada vez más estrecha y el personal bibliotecario debería asumir este rol	ULPGC

Las bibliotecas reúnen gran cantidad de datos que pueden ser de utilidad para el análisis de cuestiones como los hábitos de lectura y estudio, el uso de la información académica, etc. y, a su vez, pueden beneficiarse para su gestión de los resultados de los análisis de esos datos	ULL
Preservación de los datos de investigación y uso de la información	UPC
Hacia la localización, selección, conservación y difusión de los datos	UPSA

**Tabla VI. Líneas de trabajo futuro en el ámbito del análisis masivo de datos en bibliotecas.**

En último lugar, realizaron una importante reflexión sobre las posibles mejoras y transformación de la tecnología de análisis de grandes volúmenes de datos, llegando a resultados muy interesantes (Tabla VII). Entre ellos podemos citar: la necesidad de encontrar estándares que faciliten la interoperabilidad de los datos y garanticen la creación de procedimientos de trabajo comunes; mejorar los programas, herramientas y algoritmos que trabajan durante las diferentes etapas de la cadena de valor del *big data*; o trabajar específicamente sobre algunas de las características del análisis de datos masivos como la veracidad y la visualización.

Áreas de mejora de esta tecnología	Biblioteca universitaria
Mejora del conocimiento de todos los aspectos de la biblioteca a partir de datos objetivos	UPV
Visualización y posibilidad de reutilización	UJI
La veracidad y disponibilidad	UOC
La adopción de estándares que permitan una mejor interoperabilidad entre los datos	ULL
En bibliotecas, la gestión de necesidades de los usuarios, usos y patrones de comportamiento	UPC
Ordenadores cada vez más potentes, programas, algoritmos, etc.	UPSA

**Tabla VII. Áreas de mejora de la tecnología *big data*.**

#### 4 CONCLUSIONES

Tomando como referencia el objetivo general del que partíamos, analizar el grado de conocimiento y aplicación de la tecnología de *big data* en las bibliotecas universitarias, a partir de los resultados obtenidos podemos resaltar las siguientes conclusiones. Previamente diremos que la aplicación del *big data* en las bibliotecas universitarias se encuentra en una fase embrionaria.

En primer lugar, el conocimiento del término *big data* en líneas generales es el correcto: el empleo de gran cantidad de datos generados por las actividades humanas que no se pueden procesar con las tecnologías tradicionales. En lo referente a sus principales características, están de acuerdo en las tres “V” principales, volumen, velocidad y variedad, aunque resaltan mínimamente el valor de los datos sobre las dos últimas características reseñadas. Igualmente, existe una preferencia por las fuentes de datos internas frente a las externas.

En cuanto al procesamiento de los datos reconocen más importante el análisis de los datos que la generación y almacenamiento de los mismos. Sin embargo, es interesante que consideren que en ausencia de un profesional propio sean los bibliotecarios los que más deben adaptarse al cambio y desarrollar este rol. Sin poner en duda la actualización que deben realizar los profesionales de la información creemos que su labor fundamental debe realizarse en el estado inicial del proceso, es decir, en la descripción y representación de la información, que es el eje principal de la labor documental. Es muy interesante, que entre las competencias a desarrollar en esta nueva tarea sea la ética de trabajo en lo relativo a la explotación de los datos sensibles. Esta apreciación pone en valor la sensibilidad del bibliotecario en el manejo de los datos privados siendo conscientes que la ausencia de la misma en el tratamiento de los datos masivos es una de las principales críticas que recibe esta tecnología.

Por otro lado, es elevado el porcentaje de bibliotecarios universitarios que creen en el valor que les puedes aportar en sus centros la utilización del *big data* fundamentalmente en tres áreas: la atención al usuario, la optimización de la institución y el apoyo a la investigación. A pesar de ello, la aplicación real del *big data* en este tipo de instituciones actualmente es escasa debido primordialmente a la priorización de otras áreas o servicios y la ausencia de la tecnología adecuada. En los casos en los que se ha implantado la tecnología de los datos masivos se ha optado por el uso de programas comerciales o de código abierto.

A tenor de sus apreciaciones, el futuro “bibliotecario de datos” deberá ser un profesional que además de las tareas tradicionales de gestión y de preservación deberá sumársele la de reutilización de los datos. Esta aplicación revertirá en una atención al usuario más personalizada y en la creación de nuevos servicios más especializados. Finalmente son interesantes las reflexiones alcanzadas sobre las mejoras en el uso de los macrodatos pudiendo citar la aplicación de estándares y desarrollo de programas entre otros. Nuevamente la veracidad de los datos se considera un valor al alza.

A partir de las conclusiones presentadas anteriormente, observamos como existen posibles mejoras en cuanto a la implantación de *big data* en las bibliotecas universitarias lo que permite establecer algunas pautas a seguir para mejorar dichos aspectos, favoreciendo la utilización del análisis de macrodatos en dichos organismos para generar valor y servicios. Entre estas recomendaciones pueden recopilarse las siguientes:

- Establecimiento y definición clara de un perfil profesional con un conjunto de competencias determinadas que permita desempeñar de manera correcta todas las funciones necesarias en la aplicación de este tipo de tecnología.
- Formación específica para el personal bibliotecario sobre los conceptos y tecnología del *big data* haciendo especial énfasis en los beneficios que podría aportar este fenómeno a su entorno de trabajo.
- Aplicar dicha tecnología en mejorar los servicios de la biblioteca personalizándolos según el perfil del usuario y facilitar la optimización de la colección.
- Puesto que el análisis de los datos masivos es costoso, se deberían ofrecer subvenciones a aquellas universidades que implementen esta tecnología para favorecer el progreso tecnológico y mejorar los métodos de educación e investigación actual.
- Reutilización de los datos de los usuarios y de la biblioteca para la realización de proyectos por parte de investigadores y docentes.
- Generación de aplicaciones basadas en datos, donde los usuarios puedan interactuar con las investigaciones de una o varias bibliotecas universitarias.

Entre las futuras líneas de investigación, ya que todavía la aplicación de esta técnica se encuentra en un estado incipientemente, se debería profundizar en el uso y función de los macrodatos en el caso de las bibliotecas universitarias. En este sentido, se debería indagar en cada uno de los procesos que componen la gestión de los datos: explotación, almacenamiento y producción.

En conclusión, aunque se vislumbra un futuro prometedor de la aplicación de los datos masivos en las bibliotecas universitarias hoy en día a tenor de los resultados obtenidos podemos decir que los conocimientos teóricos son muchos pero que las aplicaciones prácticas son pocas.

## 5 BIBLIOGRAFÍA

- AL-BARASHDI, H. y AL-KAROUSHI, R. Big Data in academic libraries: literature review and future research directions [en línea]. *Journal of Information Studies & Technology (JIS&T)*, vol. 2018, nº 2. Disponible en: <https://doi.org/10.5339/jist.2018.13>.
- ALONSO, J. y VÁZQUEZ, M. Big Data: la próxima “gran cosa” en la gestión de la información [en línea]. *BiD: textos universitaris de biblioteconomia i documentació*, 2016, nº 36. Disponible en: <http://dx.doi.org/10.1344/BiD2016.36.2>.
- DE MAURO, A.; GRECO, M. y GRIMALDI, M. A formal definition of Big Data based on its essential features. *Library Review*, 2016, vol. 65, nº 3, p. 122-135. Disponible en: <http://doi.org/10.1108/LR-06-2015-0061>.
- ERWAY, R. y RINEHART, A. *If you build it, will they found?: making research data management sustainable* [en línea]. Ohio: OCLC, 2016. Disponible en: <https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016-a4.pdf> [Consulta: 22 de julio de 2019].
- FERRER-SAPENA, A. y SÁNCHEZ-PÉREZ, E.A. Open data, big data: ¿hacia dónde nos dirigimos? [en línea]. *Anuario ThinkEPI*, 2013, vol. 7, p. 150-156. Disponible en: <https://riunet.upv.es/handle/10251/38086> [Consulta: 22 de julio de 2019].
- GARCÍA-ALSINA, M. *Big Data: gestión y explotación de grandes volúmenes de datos*. Barcelona: Editorial UOC, 2017.
- GÓMEZ, J.A. La biblioteca universitaria. En: ORERA, L. (coord.), *Manual de biblioteconomía*. Madrid: Síntesis, 1996, p. 363-378.

- HARPER, L.M. y OLTSMANN, S.M. Big Data's impact on privacy for librarians and information professionals [en línea]. *Bulletin of the Association for Information Science and Technology*, 2017, vol. 43, nº 4, p. 19-23. Disponible en: <https://doi.org/10.1002/bul2.2017.1720430406>.
- HERNÁNDEZ PÉREZ, T. En la era de la web de los datos: primero datos abiertos, después datos masivos. *El profesional de la información*, 2016, vol. 25, nº 4, p. 517-525. Disponible en: <http://doi.org/10.3145/epi.2016.jul.01>.
- JUN, L.; MING, L.; GUOWEI, D. y SHANYONG, W. Big data application framework and its feasibility analysis in library [en línea]. *Information Discovery and Delivery*, 2017, vol. 45, nº 4, p. 161-168. Disponible en: <https://doi.org/10.1108/IDD-03-2017-0024>.
- GOLUB, K. y HANSSON, J. Big Data in library and information science: a brief overview of some important problem areas [en línea]. *Journal of Universal Computer Science*, 2017, vol. 23, nº 11, p. 1098-1108. Disponible en: <http://dx.doi.org/10.3217/jucs-023-11-1098>.
- RAMOS-SIMÓN, L.F. El uso de las licencias libres en los datos públicos abiertos [en línea]. *Revista Española de Documentación Científica*, 2017, vol. 40, nº 3, e179. Disponible en: <http://dx.doi.org/10.3989/redc.2017.3.1376>.
- RANI, B.R., y RAO, S.S. Big Data and Academic Libraries. Disponible en: <http://km.ptar.uitm.edu.my/documents/10180/1099975/Big+data+and+Academic+Libraries.pdf/74bbc20c-cfc0-4cae-ba4b-ead791b8f406> [Consulta: 22 de julio de 2019].
- REBIUN. Big Data en las Universidades y Bibliotecas Universitarias Españolas. Resumen de la encuesta sobre Big Data [en línea]. Disponible en: <http://repositori.uji.es/xmlui/handle/10234/163385> [Consulta: 22 de julio de 2019].
- SERRANO-COBOS, J. Big data y not so big data [en línea]. *Anuario ThinkEPI*, 2013, vol. 7, p. 161-163. Disponible en: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/30353/15949> [Consulta: 22 de julio de 2019].
- ZHAN, M. y WIDÉN, G. Public libraries: roles in Big Data [en línea]. *The Electronic Library*, 2017, vol. 36, nº 1. Disponible en: <https://doi.org/10.1108/el-06-2016-0134>.
- ZHAN, M. y WIDÉN, G. Understanding big data in librarianship [en línea]. *Journal of Librarianship and Information Science*, 2017, vol. 51, nº 2, p. 561-576. Disponible en: <https://doi.org/10.1177/0961000617742451>.
- ZETTERLUND, B. Big Data and libraries: getting the most from your library data [en línea]. Disponible en: <https://www.axiell.co.uk/getting-the-most-from-your-library-data/> [Consulta: 22 de julio de 2019].

## ANEXO I

## Cuestionario remitido a las bibliotecas universitarias

1. Presentación.
2. Datos de interés:
  - 2.1. Nombre de la biblioteca (especificar). \*
  - 2.2. ¿Cuál es su rol como biblioteca? \*
    - Biblioteca central.
    - Biblioteca de facultad.
    - Biblioteca de centro de investigación.
    - NS/NC.
    - Otro (especificar).
  - 2.3. ¿A qué ámbito científico pertenecen? \*
    - Ciencias naturales.
    - Ciencias sociales.
    - Humanidades.
    - NS/NC.
    - Otro (especificar).
3. Concepto de Big Data:
  - 3.1. ¿Conocen el término Big Data? \*
    - Sí.
    - No (Se avanza al sexto apartado).
    - NS/NC.
  - 3.2. ¿Qué entienden por Big Data? \*
  - 3.3. ¿Cuál/es de sus características consideran más importantes? \*
    - Volumen.
    - Variedad.
    - Velocidad.
    - Veracidad.
    - Valor.
    - Visualización.
    - Verificación.
    - Variabilidad.
    - Viabilidad.
    - NS/NC.
    - Otra (especificar).
  - 3.4. ¿Consideran que el empleo del Big Data en el ámbito bibliotecario es importante y/o beneficioso? \*
    - Sí.
    - No.
    - NS/NC.
  - 3.5. ¿Quiénes deberían ser los expertos en Big Data? \*
    - Bibliotecarios.
    - Informáticos.
    - Estadísticos.
    - Matemáticos.
    - NS/NC.
    - Otros (especificar).
  - 3.6. ¿Consideran que existe ya algún perfil especializado en Big Data? \*
    - Sí.
    - No.
    - NS/NC.
  - 3.7. ¿Qué competencias consideran que deben tener estos nuevos perfiles? \*
4. Big Data en su institución:
  - 4.1. En su institución, ¿cuentan con algún tipo de implementación de tecnologías Big Data? \*
    - Sí (salta la tercera pregunta).
    - No (salta las dos siguientes preguntas).

- NS/NC.
- 4.2. ¿Con que dificultades se encontraron al implementarla?
- 4.3. ¿Qué tipo de tecnologías Big Data utilizan? \*
- 4.4. ¿Por qué creen que no lo han implementado?
- 4.5. ¿En qué creen que mejora o podría mejorar su institución mediante la aplicación de Big Data? \*
- 4.6. ¿Cuáles deberían ser las principales fuentes de datos? \*
  - Fuentes internas, donde los datos se generen por la propia institución.
  - Fuentes externas, donde los datos son generados por otras instituciones.
  - NS/NC.
- 4.7. ¿Qué fase de la cadena de valor consideran más importante? \*
  - Generación de datos.
  - Adquisición de datos.
  - Almacenamiento de datos.
  - Análisis de datos.
  - NS/NC.
- 5. Perspectivas de futuro:
  - 5.1. ¿Hacia donde consideran que debe avanzar la relación entre bibliotecas y gestión de grandes volúmenes de datos?
  - 5.2. ¿Qué áreas de mejora observan en la tecnología Big Data?
- 6. Despedida:

Muchas gracias por su colaboración al responder este cuestionario.