



La gestión de datos de investigación en el horizonte de las bibliotecas universitarias y de investigación

Julio Alonso-Arévalo¹

Recibido: 2 de diciembre de 2019 / Aceptado: 3 de enero de 2019

Resumen: La sociedad TIC necesita, hace uso y proporciona de una enorme cantidad de datos; procesarlos, entenderlos y transformarlos en decisiones de valor es el reto del análisis de los llamados "big data", un término utilizado para referirse a la explosión de una ingente cantidad y diversidad de datos digitales de alta frecuencia, que se están convirtiendo en un elemento esencial para la competencia, y que según la mayoría de los expertos en el tema serán clave para el crecimiento de la productividad, la innovación y la posibilidad de generar el suficiente excedente para la sostenibilidad de la sociedad. La gestión de los datos de investigación se plantea como uno de los grandes retos que han de asumir durante los próximos años las bibliotecas científicas y de investigación. Ya la mitad de las universidades estadounidenses tienen un plan de trabajo sobre esta cuestión, todos los informes de tendencias están de acuerdo en que la Gestión de Datos de Investigación (RDM) será una de las cuestiones prioritarias y de futuro que deberán asumir las bibliotecas de investigación. En este artículo se hace un análisis del estado de la cuestión sobre la gestión de datos de investigación, normativa, repositorios de datos, prácticas y políticas que están desarrollando las bibliotecas en torno al tema.

Palabras clave: Bibliotecas universitarias, Research data management (RDM), Big Data, Tendencias, Futuro. Gestión de datos de investigación, Acceso abierto

[en] Research data management (RDM) on the horizon of academics and research libraries

Abstract: The ICT society needs, makes use of and provides an enormous amount of data; processing them, understanding them and transforming them into value decisions is the challenge of the analysis of the so-called "big data", a term used to refer to the explosion of an enormous amount and diversity of high frequency digital data, which are becoming an essential element for competition, and that according to most experts in the field will be key to productivity growth, innovation and the possibility of generating sufficient surplus for the sustainability of society. The management of research data is one of the major challenges facing scientific and research libraries in the coming years. Already half of American universities have a work plan on this issue, all trend reports agree that Research Data Management (RDM) will be one of the priority and future issues to be taken up by research libraries. This article analyses the state of research data management, regulations, data repositories, practices and policies that libraries are developing around the topic.

Keywords: : Academic libraries, Research Data management (RDM), Big Data, Trends, Future. Open Access

¹ Jefe de la Biblioteca de Traducción y Documentación. Universidad de Salamanca (España)
E-mail: alar@usal.es

Sumario: 1. Introducción 2. ¿Cuál es la importancia de la gestión de datos? 3. ¿Cómo gestionar los datos? La gestión de datos de investigación (RDM) 4. Repositorios de gestión de datos de investigación 5. Aplicaciones 6. Papel de la biblioteca y los bibliotecarios en la gestión de datos. 7. Conclusiones 8. Referencias bibliográficas

Cómo citar: Alonso-Arévalo, Julio (2019). La gestión de datos de investigación en el horizonte de las bibliotecas universitarias y de investigación. *Cuadernos de Documentación Multimedia*, 30, 75-88.

1. Introducción

El aumento del volumen y organización de la información capturada por las empresas y organizaciones, el incremento de los multimedia, las redes sociales y la “Internet de las cosas” están impulsando un crecimiento exponencial de los datos en el futuro. Datos de registros de llamadas, transacciones de banca móvil, contenidos generados por el usuario de internet, tales como blogs y tweets, búsquedas en línea, imágenes de satélite, etc. es información procesable que requiere el uso de técnicas computacionales para dar a conocer las tendencias y patrones dentro de éstos extremadamente grandes conjuntos de datos. Procesarlos, entenderlos y transformarlos en decisiones de valor es el reto de los llamados «big data».

Según estima idC los datos se incrementan un 50% al año, o sea que se duplican cada dos años. En áreas tan variadas como la ciencia y los deportes, la publicidad y la salud pública se ha producido un salto hacia el descubrimiento y la toma de decisiones a partir de los datos. Según Stanto, la Ciencia de los Datos se refiere a un área emergente de trabajo que se ocupa de la recogida, preparación, análisis, visualización, administración y conservación de grandes colecciones de información. Aunque el nombre de Datos Científicos parece ser más propio de áreas de conocimiento tales como bases de datos y la informática, incluye muchos tipos diferentes de habilidades – incluyendo habilidades no-matemáticas -(Stanton, 2013). Casi todos los analistas consideran “Big Data” como una de las tendencias de futuro que tendrán que tener en cuenta la mayoría de las empresas e instituciones. (Wanner, 2015)

“Big Data” es un término utilizado para referirse a la explosión de una ingente cantidad y diversidad de datos digitales de alta frecuencia, que se están convirtiendo en un elemento esencial para la competencia, y en un futuro inmediato serán clave para el crecimiento de la productividad, la innovación y la posibilidad de generar el suficiente excedente para la sostenibilidad de la sociedad (Erway & Rinehart, 2016). Por ello, la gestión de los datos de investigación (RDM) se plantea como uno de los grandes retos que han de asumir durante los próximos años las universidades y las bibliotecas científicas y de investigación.

La agencia Gartner proporciona una descripción del término en la siguiente frase: “*Big Data se refiere al volumen, variedad y velocidad de datos estructurados y no estructurados que se vierten a través de redes en los*

procesadores y dispositivos de almacenamiento, así como la conversión de dichos datos para el asesoramiento empresarial”.

Estos elementos se pueden dividir en tres categorías distintas: volumen, variedad y velocidad.

— Volumen (terabytes, petabytes y exabytes, eventualmente): La cantidad cada vez mayor de datos generados por los seres humanos y las máquinas está imponiendo un reto a los sistemas informáticos, que están trabajando con el propósito de almacenar, proteger y poner a disposición de la sociedad toda esta información para su uso futuro.

— Variedad: Big Data es también el creciente número de tipos de datos que deben ser manejados de manera diferente, teniendo en cuenta datos básicos como pueden ser simples correos electrónicos, registros de datos, registros de tarjetas de crédito, estudios científicos, registros de salud, datos financieros y multimedia: fotos, presentaciones gráficas, música, audio y vídeo.

— Velocidad: Se trata de la velocidad a la que estos datos se mueven a partir de criterios de valoración en el procesamiento y almacenamiento.

2. ¿Cuál es la importancia de la gestión de datos?

Un informe del Foro Económico Mundial declaró que los datos constituyen una nueva clase de activo económico, como la moneda o el oro. Para la revista Forbes, el Big Data ha sido la principal tendencia tecnológica de los últimos años que se mantendrá los próximos años. Para la IDC, hasta 2015 su crecimiento será siete veces superior al de la media de todo el sector de las TIC (Parinov & Antonova, 2016) .

Por ello, los datos son considerados como “la nueva materia prima del siglo 21”, sin embargo, en la actualidad los datos de investigación quedan encerrados en las oficinas de los investigadores, o en las instituciones y empresas que los crearon. Sin embargo, la mayoría de las instituciones académicas, organizaciones de investigación, las principales editoriales requieren que los datos se presenten adecuadamente para poder utilizarlos, que estén disponibles para cualquier propósito y puedan reutilizarse para nuevas investigaciones.

Según un informe de MGI y la Oficina de Tecnología de Negocios de McKinsey (Mannyica, Michael, Brown, & Bughing, 2011) la cantidad de información que genera cualquier actividad pública o privada proporciona grandes conjuntos de datos, y el análisis de estos se ha convertido en una de las bases clave para la competencia en un futuro inmediato que sustentará las nuevas oleadas de crecimiento, de productividad, innovación y excedente del consumidor.

Un requisito indispensable para todo ello es que estos datos este accesible para todos. Los datos se definen como abiertos cuando se pueden utilizar libremente, modificar y compartir por cualquiera para cualquier propósito, incluyendo la creación de un bien común en el que cualquiera puede participar. Los datos abiertos son de utilidad para millones de personas en todo el mundo, investigadores, empresas y ciudadanos. Pero para que puedan ser realmente útiles es necesario crear aplicaciones, poner en marcha nuevos productos y servicios que mejoren la transparencia y la apertura.

En los últimos años, varias organizaciones científicas nacionales e internacionales han emitido declaraciones y políticas que subrayan la necesidad de un inmediato archivo de los datos, y algunos organismos de financiación han comenzado a exigir que los datos procedentes de investigaciones que se financian con recursos públicos sean depositados en un archivo abierto. Estas declaraciones de las principales agencias de financiación de la investigación demuestran que la ética del intercambio de datos es esencial para maximizar el impacto y los beneficios de la investigación (Committee, 2014). Ya que, la experiencia ha demostrado que la dispersión de los datos aumenta el coste de procesamiento y que la preservación de los datos disminuye la inversión cuando se dispone de depósitos bien conformados.

A su vez, las más importantes empresas del sector de la información científica han iniciado plataformas para la compilación y difusión de datos de investigación. De este modo Thompson Reuters creó la plataforma Data Citation Index, que permite acceder a un conjunto de datos a través de temas y regiones, proporcionando una imagen completa de la producción de investigación para comprender los datos en su contexto y maximizar los esfuerzos de investigación. En este mismo sentido, otra de las grandes de la edición científica como Elsevier contribuye al tema datos de investigación con dos proyectos *Mendeley Data*, un repositorio de datos de investigación y *DataSearch*, un motor de búsqueda que permite buscar conjuntos de datos de investigación a través de numerosos dominios y tipos (imágenes, documentos, datos brutos. etc.) a partir de un conjunto de repositorios de datos. Por otra parte, seis de las grandes bibliotecas universitarias y de investigación estadounidenses, han creado una red para desarrollar el proyecto “*Data Curation Network*”, con el objetivo de mejorar el apoyo al investigador. El proyecto tiene como base preparar los datos digitales de investigación para el acceso abierto y la reutilización. Lo que permitirá a las bibliotecas universitarias trabajar colectivamente y de manera más eficaz, con una variedad más amplia de tipos de datos (por ejemplo, disciplina, formato de archivo, etc.) más allá de lo que podría ofrecer una institución por sí sola. El desarrollo tiene como objetivo investigar y definir los flujos de trabajo y mecanismos de curación de datos a través de reuniones de un equipo profesional estructurado, para definir las pautas de trabajo a partir de otras experiencias y conocimientos previos adquiridos por cada una de las instituciones participantes. Para ello se han planteado tres acciones:

1. Establecer parámetros y controlar el esfuerzo (el costo, tiempo, experiencia) en cada una de las seis instituciones participantes (Universidad de Minnesota, Cornell University, Penn State University, Illinois University, Michigan University y Washington University St. Louis)

2. Solicitar información a los investigadores para comprender en qué medida los servicios de curación de datos se adaptan a las necesidades del flujo de trabajo de investigación y gestión de datos, a través de las actividades de participación informales desarrolladas en paralelo en cada uno de los campus.

3. Desarrollar un modelo para el intercambio de datos entre las instituciones con el objetivo de proporcionar servicios de custodia de información en el que se tenga en cuenta la dotación de personal, costos, conjuntos de habilidades, y la demanda necesaria para su implementación.

El resultado de este proyecto pretende establecer un modelo de gestión y curación de datos que proporcionará una serie de conocimientos sobre cómo puede ser implementado, evaluado, y sostenido por un grupo de instituciones académicas, proyecto que podrá ser de interés para otras experiencias futuras.

El intercambio de datos también permite a los científicos utilizar, analizar y reproducir los resultados de los demás. Aunque hay muchos más beneficios derivados del intercambio de datos que van más allá de la mera posibilidad de replicación. Fienberg (*Guide to Social Science Data Preparation and Archiving: Introduction.*, 2012) sostiene que el intercambio de datos:

- Refuerza la investigación científica abierta. Cuando los datos están ampliamente disponibles, las posibilidades de auto-corrección de la ciencia son más eficaces.
- Alienta la diversidad de análisis y opiniones. Si los investigadores tienen acceso a los mismos datos pueden más fácilmente desafiar los análisis y conclusiones de los otros.
- Promueve la investigación y permite la comprobación de los métodos nuevos o alternativos. Existen muchos ejemplos de datos que se utilizan de una manera que los investigadores originales no habían previsto.
- Mejora métodos de recogida y medición de datos a través del escrutinio de los demás. La puesta a disposición del público permite a la comunidad científica llegar a un consenso sobre los métodos.
- Reduce los costes evitando los esfuerzos de recolección de datos duplicados. Algunos conjuntos de datos estándar tales como encuestas sociales y electorales producen miles de documentos que no podrían haber sido publicados si los autores tuvieran que recoger sus propios datos.
- El archivado pone a disposición de todos los datos para que no se tengan que invertir recursos adicionales para reunir esencialmente la misma información.
- Proporciona un recurso importante para la formación en investigación. Los datos secundarios son extremadamente valiosos para los estudiantes, que de ese modo tienen acceso a datos de alta calidad como modelo para su propio trabajo.
- El archivado inmediato permite a un investigador mejorar el impacto (y, ciertamente, la visibilidad) de un proyecto.

3. ¿Cómo gestionar los datos? La Gestión de Datos de Investigación (RDM)

La Gestión de Datos de Investigación (RDM) es un proceso diseñado para gestionar y difundir conjuntos de datos de alta calidad, que cumplan con los requisitos académicos, legales y éticos establecidos.

Varias instituciones, asociaciones, y organizaciones han empezado a dar pautas para la adecuada gestión de datos de investigación. Así grupo de trabajo LIBER E-Ciencia, creado para investigar el papel que las bibliotecas deben desempeñar en el ámbito de la e-Ciencia., decidió concentrarse en los datos de la investigación, ya que se consideró que esta era la necesidad más urgente de la *e-ciencia* para la comunidad de investigación. Por ello publicó un informe sobre la gestión de datos de investigación titulado "*Ten recommendations for libraries to get started with research data management*" (Steven Bell, Lorcan Dempsey, & Barbara Fister, 2015), con 10 recomendaciones para que las bibliotecas inicien servicios de gestión de datos de investigación (RDM):

1. Iniciar propuestas sobre Gestión de Datos de Investigación (RDM) que incluyan planes de gestión de datos para la solicitud de subvenciones, asesoramiento sobre derechos de propiedad intelectual y material informativo. Ayudar a los profesores con los planes de gestión de datos y la integración de la gestión de datos en los planes de estudios.

2. Participar en el desarrollo de estándares de metadatos para proporcionar servicios sobre la gestión de datos de investigación.

3. Difundir y desarrollar las habilidades necesarias para la gestión de datos entre el personal profesional.

4. Participar activamente en la investigación para el desarrollo de políticas de datos institucionales, incluyendo los planes de recursos. Fomentar y adoptar políticas de apertura en todo el ciclo de vida de los datos de investigación.

5. Mantener contactos y colaborar con los investigadores, grupos de investigación, archivos de datos y centros de datos para fomentar una infraestructura interoperable para el acceso a datos, descubrimiento y difusión de los datos de investigación.

6. Apoyar todo el ciclo de vida para los datos de investigación, proporcionando servicios para el almacenamiento, el descubrimiento y el acceso permanente.

7. Promover la investigación sobre datos mediante la aplicación de identificadores persistentes a los datos de investigación.

8. Proporcionar un catálogo de datos institucional o repositorio de datos, dependiendo de la infraestructura disponible.

9. Participa en la práctica de gestión de datos en la disciplina de trabajo específica.

10. Oferta o mediar en torno a cuestiones como el almacenamiento seguro de los datos de la investigación dinámica y estática en cooperación con las unidades que gestionan TICs a nivel institucional y / o buscar la explotación de los servicios apropiados. en la nube

ACRL también publicó un libro blanco sobre la gestión de datos de investigación. (Tenopir, Birch, & Allard, 2012) que propone algunos puntos previos que es necesario desarrollar:

- La creación de una guía web para ayudar a localizar los datos ofrecidos puede ser un buen comienzo para extender las prácticas y la conciencia sobre la gestión de datos de investigación (GDI) en el nuevo entorno.

- La gestión de datos de investigación (GDI) requiere de la colaboración con otras unidades en el campus como la oficina de apoyo a la investigación. Esta colaboración es una oportunidad excelente para que las bibliotecas asuman el liderazgo en el proceso de creación de conocimiento, lo que será de gran ayuda para la mejor visión, misión y valoración de la biblioteca por parte de la comunidad universitaria.

- La reasignación al proceso de gestión de datos de personal existente de la biblioteca es la táctica más común que están desarrollando las bibliotecas para ofrecer servicios de GDI. Este enfoque también debe ser compatible con el desarrollo profesional para el personal, ya que facilitará la adquisición de los conocimientos necesarios para proporcionar una gama completa de servicios de gestión de datos

- De las bibliotecas que proporcionan GDI, la mayoría han reasignado, o tiene intención de volver a asignar personal existente. Esto que puede verse como una debilidad financiera, puede ser un potencial como un medio para el desarrollo de un argumento que permita obtener fondos adicionales para nuevas propuestas que tengan relación con la GDI.

- Las bibliotecas están ofreciendo conferencias o talleres para capacitar a su personal en gestión de datos de investigación (GDI)

- Las bibliotecas necesitan algún tipo de apoyo institucional para capacitar a sus bibliotecarios en esa área de desarrollo profesional, y es importante para las organizaciones profesionales seguir proporcionando esta formación. Puede haber una oportunidad formativa para aquellas bibliotecas que inicien antes estos procesos para crear una relación de tutoría con compañeros u otros asociados, para ayudar en la formación de otros profesionales a través de sus propias experiencias.

Para Johnsson y Åhlfeldt, llegados a este punto, cabe preguntarse ¿Qué aspectos se deben tener en cuenta en este proceso? (Johnsson & Åhlfeldt, 2015):

- ¿Quién posee los datos?
- ¿Qué requisitos son impuestos por otros?
- ¿Qué datos deben conservarse?
- ¿Por cuánto tiempo deben conservarse estos datos?
- ¿Cómo deben ser preservados los datos digitales?
- ¿Existen consideraciones éticas?
- ¿Cómo se accede a los datos?
- ¿Cómo deben ofrecerse los datos?
- ¿Cómo se manejarán los costos?
- ¿Cuáles son las alternativas para la gestión de datos locales?

Todo este proceso se desarrolla a través de dos líneas 1. La preservación a largo plazo de los conjuntos de datos mediante sistemas de almacenamiento 2. Compartición y reutilización de los conjuntos de datos para la investigación y otros fines en la sociedad en general

De este modo, los investigadores deben planificar un proyecto eventual de archivo y difusión de los datos antes de que los datos lleguen a existir. El archivo de datos debe contemplarse desde el inicio de cualquier proyecto, estableciendo un calendario para depositar los diversos productos en el transcurso del ciclo de vida de los datos, abarcando la creación y preservación de metadatos precisos, que garanticen la facilidad de uso de los propios datos de investigación. Tales prácticas incorporarían el archivado y preservación como parte del método de investigación.

El proceso implica desarrollar un plan para hacer frente a las consideraciones de archivos que entran en juego en todas las etapas del ciclo de vida de los datos. (Jacobs y Humphrey, 2004). Teniendo en cuenta todas las etapas y procesos de datos desde la fase de identificación, pertinencia y recogida, hasta la fase de preparación de los datos para su depósito en los archivos o repositorios de datos. Es decir, requiere de un plan que abarque todo el ciclo vital de los datos. (Ver figura)

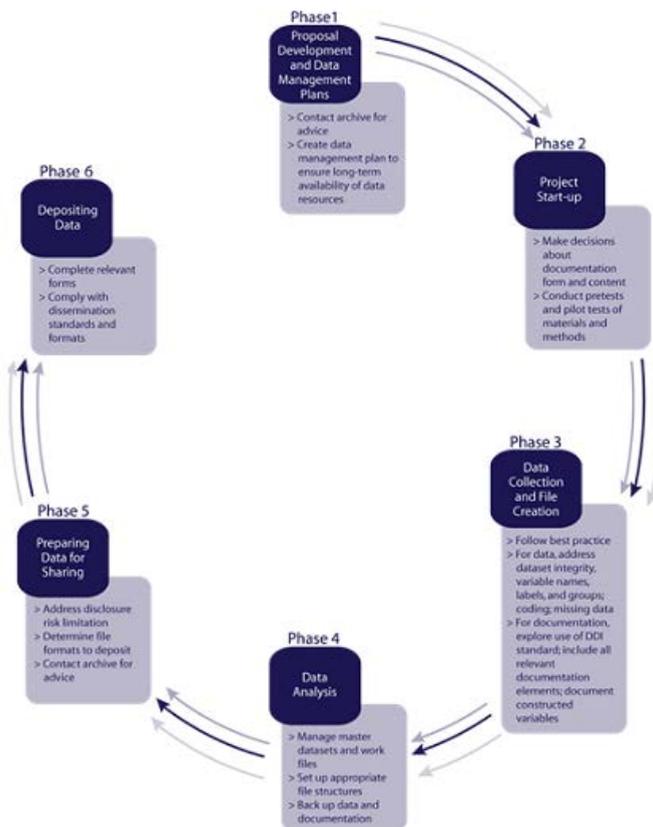


Fig. 1. Intervención en el ciclo vital de los datos. Jacobs y Humphrey, 2004

4. Repositorios de gestión de datos de investigación

Anteriormente afirmamos que la variedad de datos de investigación es muy amplia y heterogénea, datos de patrones de comportamiento, datos matemáticos, datos textuales, etc. De ahí que los repositorios de datos varían mucho en cuanto a sus contenidos, objetivos, métodos y políticas de acceso. Dependiendo de la materia o disciplina de investigación, los datos generalmente pueden ser depositados en más de un centro de datos o repositorio. Cada depósito tiene sus propios requisitos o especificaciones respecto a los datos que ofrece, sobre la base de la investigación objeto o de dominio, metadatos, formato de archivo y / o estructura de datos, y en función de los tipos de datos y la naturaleza de las políticas de reutilización y de acceso. Un aspecto importante es que estos datos puedan reutilizarse y/o remezclarse.

Herramientas para la recopilación de Ciencia Abierta y Datos de Investigación:

Generales

- [Dryad Digital Repository](#), a resource for making the data from scientific publications discoverable, reusable, and citable
- [Elsevier DataSearch](#). Search for research data across domains and types, from many domain-specific, cross-domain and institutional data repositories.
- [Mendeley Data](#) Put your research data online today
- [protocols.io](#) Discover & share science protocol knowledge
- [re3data.org](#) the global registry of research data repositories
- [Repositiv](#) Data discovery platform for human genomic data
- [OBF](#) Promote the practice and philosophy of Open Source software development
- [Figshare](#) Manage your research in the cloud
- [LabScribbles](#) Blog about real-time open access science
- [Zenodo](#) Share all research outputs from across all fields of science
- [OSF](#) Platform that simplifies scientific collaboration

Por disciplinas

Ver Springer Nature [list of trusted data repositories - Springer Nature](#)

5. Aplicaciones

En cuanto al cómo aún no hay una definición específica, ya que se trata de un asunto complejo debido a la gran cantidad y variedad de datos, pero ya existe alguna herramienta como Dataverse, una aplicación web de código abierto desarrollada por la Universidad de Harvard que permite compartir, preservar, citar, explorar y analizar datos de investigación. El programa facilita la toma de datos y los pone a disposición de los demás, permitiendo replicar otros trabajos de investigación. Un repositorio Dataverse aloja varios *dataverses*. Cada dataverse contiene un dataset (grupos de datos) u otros dataverses, y cada conjunto de datos contiene metadatos descriptivos y archivos de datos (incluyendo la documentación

y el código que acompañan a los datos). Dataverse normaliza la cita de los conjuntos de datos para que sea más fácil para los investigadores publicar sus datos y obtener un mejor reconocimiento de su trabajo. Cuando se crea un conjunto de datos en Dataverse, se genera la citación y se presenta de forma automática como un marco único de código abierto y repositorio de datos de investigación, lo que hace que los datos científicos sean lo más accesibles, reutilizables, y abiertos posible

6. Papel de la biblioteca y los bibliotecarios en la gestión de datos.

Una de las características más innovadoras de la biblioteca del siglo 21 tiene que ver con la toma de una postura activa frente a la gestión de datos de investigación, así lo ponen de relieve los principales informes sobre tendencias en bibliotecas universitarias y de investigación. La mayoría de informes sobre tendencias en bibliotecas de investigación coinciden en considerar GDI como una necesidad ineludible en el futuro de estas instituciones (Johnson, Adams, & Cummins, 2015), (Steven Bell, Lorcan Dempsey, & Barbara Fister, 2015), (Hassan & Gillani, 2016).

En casi todos estos informes se pone de relieve las capacidades de los profesionales de las bibliotecas como elementos clave en la GDI, ya que durante siglos los bibliotecarios hemos sido expertos en la organización de las colecciones y sabemos cómo encontrar las cosas. La transición a los formatos digitales ha traído consigo un enorme volumen de datos que necesita ser curados, como los relativos a descargas, citas, y la cobertura de los medios de comunicación (*Environmental Scan 2015*, 2015). Todo ello ha ido acrecentando las competencias y habilidades de los bibliotecarios como uno de los segmentos profesionales más capacitados para la gestión de grandes cantidades de datos. (Stang, 2016)

Sin embargo, conocer y alfabetizar sobre estas cuestiones requiere un esfuerzo de readaptación profesional para fomentar una mentalidad sobre la importancia de estos datos y la cultura de análisis, ya que se trata de la adopción de las nuevas tecnologías, ello presenta nuevos desafíos para los bibliotecarios. Con cuestiones pendientes de resolver aún tales como ¿Cómo pueden las bibliotecas desempeñar su papel para desarrollar una “mentalidad basada en datos”? Respecto a la formación queda por definir qué es la alfabetización de datos y en qué se diferencia de su contraparte la alfabetización informacional, o ¿qué puede aportar a un plan de estudios la alfabetización de datos en instituciones de educación superior? (*Big Data Now*, 2011).

Según David Lankes, (*Research Data Discover how Elsevier's supporting researcher store, share ,discover and use data*, 2016) un bibliotecario no tiene que convertirse en un programador, pero debe estar interesado en la creación de conocimiento, debe tener cierta familiaridad con el manejo de las diversas herramientas de software que pueden transformar los datos. Un bibliotecario no tiene que ser un ingeniero de bases de datos, pero debe comprender los fundamentos de las herramientas de recuperación de información. Un bibliotecario no tiene por qué ser un estadístico, pero debe tener una comprensión clara de cómo se gestionan los datos numéricos para que puedan ser adecuadamente utilizados.

Por último, un bibliotecario no tiene por qué ser un diseñador gráfico, pero tiene que reconocer las características necesarias para hacer una presentación de datos eficaces.

El análisis de las necesidades de datos a través de dominios institucionales puede requerir de la participación de la biblioteca para identificar y conectar a los investigadores con todas las unidades funcionales, tanto formales e informales para compartir, analizar, y reutilizar datos. Y los bibliotecarios pueden utilizar su experiencia para etiquetar y organizar este tipo de información, haciendo una contribución estratégica a su institución. (Bernal, Oficina Técnica de Digital, & Román-Molina, 2014) De hecho, los bibliotecarios llevamos algunos años desplegando conocimientos técnicos y expertos en promover y apoyar la gestión de repositorios y el intercambio de datos abiertos, por lo que la gestión de datos de investigación ha de formar parte de un desarrollo natural en nuestras tareas y funciones. Según MJ Tooley bibliotecaria de la University of Maryland *"Tenemos el conjunto de habilidades necesarias para organizar las cosas. Entendemos los vocabularios controlados. Entendemos las ontologías. Sabeos como organizar la información. Hemos realizado la evolución de la palabra impresa y el encabezamiento de materia a los datos como un proceso natural. Así que estamos preparados para ayudar a la gente a organizar, acceder y almacenar datos"*

Las más importantes bibliotecas universitarias estadounidenses ya están respondiendo a esta nueva misión a través de una variedad de servicios, aproximadamente la mitad de todas las bibliotecas universitarias tienen algún tipo de programa de apoyo a la gestión de datos. Los bibliotecarios también están haciendo un trabajo interesante para ayudar a gestionar esta evolución basada en los principios de apertura e interoperabilidad. Esta práctica es necesaria para la transición desde la idea tradicional que se tiene de una "infraestructura de datos" centrada en torno a las revistas especializadas o repositorios institucionales hacia un sistema más robusto centrado en los datos de investigación. La creación de este nuevo ecosistema requerirá de la evolución de muchas áreas, incluyendo el desarrollo de nuevos estándares en torno a la validación de los datos de investigación, los procedimientos para documentar la procedencia de los conjuntos de datos, y los nuevos modos de gestión y propiedad de los datos de investigación. Así los bibliotecarios se están mostrando activos en todos estos espacios.

Las bibliotecas utilizamos habitualmente identificadores persistentes de objetos digitales (DOI) para caracterizar de manera unívoca conjuntos de datos específicos utilizando servicios como [DataCite](#) y [EZID](#). Organizaciones como JISC han desarrollado [RDMRose](#), un proyecto para producir materiales de aprendizaje para el desarrollo profesional continuo (DPC) en Gestión de Datos de Investigación (GDI) para los profesionales de la información. También será de gran ayuda el documento publicado por OCLC [Role of Libraries in Data Curation](#) sobre el papel de las bibliotecas en el proceso de "curación" de datos.

7. Conclusión

La adecuada gestión de datos es fundamental para maximizar la utilidad y el valor de los datos de investigación de alta calidad y, por tanto, la investigación de excelencia. La organización y archivo de los datos también es crucial para facilitar el intercambio de datos y asegurar la sostenibilidad y la accesibilidad de los datos a largo plazo, y por lo tanto su reutilización para la ciencia futura. En este contexto es determinante la labor de la biblioteca en el apoyo a los investigadores de cara a gestionar y compartir datos a través de herramientas, así como el apoyo y orientación a los investigadores en torno al tema, la formación práctica; facilitar la citación y vinculación de los datos con publicaciones con el objetivo de proporcionar mayor visibilidad y accesibilidad de los datos y la investigación misma (Erway, Horton, Nurnberger, Otsuji, & Rushing, 2016)

Por el momento, sólo una pequeña parte de las bibliotecas universitarias de Estados Unidos y Canadá ofrecen actualmente servicios relacionados con la gestión de datos de investigación (GDI), aunque la mayoría de las bibliotecas se están planeando ofrecer algunos servicios relacionado con la cuestión en los próximos dos años.

Los desafíos son muchos; éstos incluyen el aprendizaje sobre los principios de gestión de datos de investigación (GDI), los problemas, la evaluación de las instituciones con mayores necesidades, selección y aplicación de un entorno de almacenamiento, colaboración con investigadores para transmitir la importancia del proyecto, la preparación de materiales de formación, la construcción de conocimientos técnicos entre personal de la biblioteca, y el establecimiento de directrices de metadatos. Por eso las universidades que forman a los profesionales de las bibliotecas están introduciendo la curación de datos y la Gestión de Datos de Investigación (GDI) en sus planes de estudios para preparar la próxima generación de profesionales de la información, y formando a bibliotecarios a través de cursos para la asunción de estos nuevos roles, que serán determinantes en la labor del profesional de un futuro más que inmediato.

El objetivo es hacer que el uso de estos datos pueda proporcionar nuevos y valiosos servicios o mejorar la eficiencia. El problema para alcanzar estas metas es que a medida que la cantidad de almacenamiento y procesamiento ha crecido, la complejidad de los datos y los retos de gestión son más complejos. La tarea esencial del profesional de la ciencia de datos es transformar, datos brutos desordenados en conocimiento procesable que puede ser utilizado por los tomadores de decisiones.

8. Referencias bibliográficas

Bell, S., Dempsey, L., & Fister, B. (2015). *New Roles for the Road Ahead*. Chicago: Association of College and Research Libraries. Disponible en: http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/new_roles_75th.pdf

- Bell, S., Dempsey, L., & Fister, B. (2015). *New Roles for the Road Ahead: Essays Commissioned for ACRL's 75th Anniversary*: ACRL. Disponible en: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/new_roles_75th.pdf
- Bernal, I., Oficina Técnica de Digital, C., & Román-Molina, J. (2014). Prácticas en la gestión, difusión y preservación de datos de investigación en el CSIC: Digital CSIC. Disponible en: <http://hdl.handle.net/10261/92404>
- Big Data Now*. (2011). Cambridge: O'Reilly Media. Disponible en: http://www.onmeedia.com/downloads/Big_Data_Now_Current_Perspectives_from_OReilly_Radar.pdf
- Committee, A. R. P. a. R. (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *C&RL News*(June 2014), 294-302. Disponible en: <http://crln.acrl.org/content/75/6/294.full.pdf+html>
- Environmental Scan 2015*. (2015). ACRL Research Planning and Review Committee. Disponible en: <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/EnvironmentalScan15.pdf>
- Erway, R., Horton, L., Nurnberger, A., Otsuji, R., & Rushing, A. (2016). *Building Blocks: Laying the Foundation for a Research Data Management Program*. Dublin, Ohio: OCLC. Disponible en: <http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-data-management-building-blocks-2016.pdf>
- Erway, R., & Rinehart, A. (2016). *If You Build It, Will They Fund? Making Research Data Management Sustainable* OCLC. Disponible en: <http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016-a4.pdf>
- Guide to Social Science Data Preparation and Archiving: Introduction*. (2012). ICPSR Inter-university Consortium for Political and Social Research. Disponible en: <http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- Hassan, S.-U., & Gillani, U. A. (2016). Altmetrics of "altmetrics" using Google Scholar, Twitter, Mendeley, Facebook, Google-plus, CiteULike, Blogs and Wiki. *arXiv.org*. Disponible en: <http://arxiv.org/abs/1603.07992>
- Johnson, L., Adams, S., & Cummins, M. (2015). *Horizon Report > 2015 Higher Education Edition*: NMC. Disponible en: <http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf>
- Johnsson, M., & Åhlfeldt, J. (2015). *Research Libraries and Research Data Management within the Humanities and Social Sciences* Lund Disponible en: <http://lup.lub.lu.se/record/5050462/file/5050466.pdf>
- Mannyica, J., Michael, C., Brown, B., & Bughing, J. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York: McKinsey & Company. Disponible en: <http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>

- Parinov, S., & Antonova, V. (2016). "End of Publication? Open Access and a New Scholarly Communication Technology". *arXiv*. Disponible en: <https://arxiv.org/ftp/arxiv/papers/1608/1608.05505.pdf>
- Research Data Discover how Elsevier's supporting researcher store, share, discover and use data.* (2016). Elsevier. Disponible en: https://www.elsevier.com/_data/assets/pdf_file/0006/156948/Research-Data-Brochure-2016.pdf
- Stang, T. (2016). Librarians: the new research data management experts : How growth in research data is spurring a shift in the librarian's role *Elsevier*. Disponible en: <https://www.elsevier.com/connect/librarians-the-new-research-data-management-experts>
- Stanton, J. M. (2013). *An Introduction to Data Science*. Syracuse: Syracuse University. Disponible en: https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf
- Tenopir, C., Birch, B., & Allard, S. (2012). *Academic Libraries and Research Data Services: Current Practices and Plans for the Future. An ACRL White Paper*: ACRL. Disponible en: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Wanner, A. (2015). Data literacy instruction in academic libraries: best practices for librarians. *Archival and Information Studies Student Journal*. Disponible en: <http://ojs.library.ubc.ca/index.php/seealso/article/download/186335/185482>