Servicios de Información para la Investigación Intensiva en Datos en el Entorno Académico Post-Internet

Lourdes Feria Basurto

Introducción: la evolución de Internet y el mundo de los datos

Programas de cómputo, protocolos de telecomunicaciones, algoritmos informáticos, redes de cableado submarino, servidores informáticos, antenas de microondas e incontables kilómetros de fibra óptica conforman el marco de telecomunicaciones que se conoce como Internet. La historia de esta red nació en los años cincuenta del siglo pasado con las primeras aplicaciones computacionales locales y más adelante con el Proyecto Arpanet que permitió la investigación sobre *networking* tecnológico. Vendrían luego las primeras transmisiones de mensajes vía red, el surgimiento del protocolo TCP/IP y el lanzamiento de la *World Wide Web*. Poco después se desarrollarían los primeros motores de búsqueda, dando comienzo en 1998 la investigación de lo que hoy en día es la compañía tecnológica Google. En 2001 nace la Wikipedia,

seguida de las redes sociales (Facebook en 2004 y YouTube al año siguiente) y el lanzamiento del *iPhone* en 2017 detonante de la cultura de las de aplicaciones móviles (Graham y Dutton 2019).

Más de 3.9 billones de personas están en línea, de ellos cuatro de cada cinco son habitantes de países desarrollados, en tanto que en los 47 países menos desarrollados del planeta un 80% de la población no cuenta con la posibilidad de ingresar en el mundo de la red). El acceso a la banda ancha se mantiene en crecimiento, especialmente para el caso de los teléfonos celulares; en 2007 había 4.0 suscripciones por cada 100 habitantes, en tanto que para el año 2018 se incrementó a 69.3. Otros puntos importantes tienen que ver con el hecho de que casi toda la población se beneficia de la cobertura de señal móvil 3G o superior y el acceso a Internet desde los hogares se ha incrementado del 20% en 2005 al 60% en 2018 (International Telecommunication Unit 2018, 2).

Nunca antes en la historia de la humanidad se produjeron las condiciones que actualmente se observan en la dinámica de la información. Su flujo genera grandes volúmenes de datos, fenómeno que se ha denominado *Big Data* (Datos Masivos) siendo sus principales características la abundancia, diversidad (ya que lo mismo integra fotografías que información personal, científica o incluso información poco confiable) y agilidad en el acceso.

Los datos son la evidencia del proceso científico, obtenidos de observaciones, experimentos o simulaciones (Monash University 2014) y la utilización del término "dato" surge en el siglo diecisiete:

> […] como en el plural de la palabra Latina *datum*, que significa una pequeña pieza de información, los datos pueden ser considerados como cualquier información que puede ser archivada

> en formato digital, incluyendo texto, números, imágenes, video, audio, software, algoritmos, ecuaciones, animaciones, modelos y simulaciones. (Semeler, Pinto y Rozados 2019, 772).

En cuanto a los tipos de datos Martínez-Uribe y Fernández (2015, 195) los clasifican en:

> […] textuales, numéricos geográficos, imágenes, videos, audios, modelos, fórmulas, algoritmos o cualquier combinación de los anteriores, contenidos en documentos de texto, hojas de cálculo, cuadernos de laboratorio, cuestionarios, cintas y casetes, fotos y películas, diapositivas, dispositivos, especímenes, muestras, etc.

Posteriormente ilustran esta diversidad con casos de algunas disciplinas científicas, indicando que, por ejemplo, para las ciencias políticas sus fuentes de datos podrían ser las encuestas o estudios de opinión, para la astronomía serían las mediciones de imágenes de telescopios o simuladores, para la biología marina las anotaciones puntuales a partir de la observación directa de especímenes de flora y fauna; en tanto que para las humanidades los datos podrían provenir de textos históricos o bien de grabaciones sonoras en el caso de estudios musicales (*Ídem*).

Gestión de datos de investigación y el cuarto *paradigma*

En la publicación de Hey, Tansley, y Tolle (2014), *El cuarto paradigma: descubrimiento científico intensivo en datos*, se reflexiona sobre la investigación científica y la computación intensiva de datos, proponiendo la idea de que este paradigma se presenta en el momento actual de una manera un

tanto similar a lo que ocurrió en el siglo XV con la invención de la imprenta y, de la misma forma que aquella se ha venido desarrollando y evolucionando a través de centurias se vislumbra que la obtención de conocimientos a partir de los datos reutilizables y los repositorios digitales será cada vez más común (*Idem* xviii).

En la ilustración 1 se muestra la evolución en las formas de hacer ciencia, observándose que en sus inicios se concretaba a la descripción de fenómenos naturales; en una segunda etapa se fortaleció con la utilización de modelos y más adelante, con el surgimiento de las computadoras se trabajó con herramientas automatizadas como los simuladores llegando al momento actual en que el paradigma de la e-ciencia abre la posibilidad de una investigación intensiva en datos gracias a que estos pueden ser recabados, procesados y analizados mediante instrumentos, software y otros medios (*Idem*).

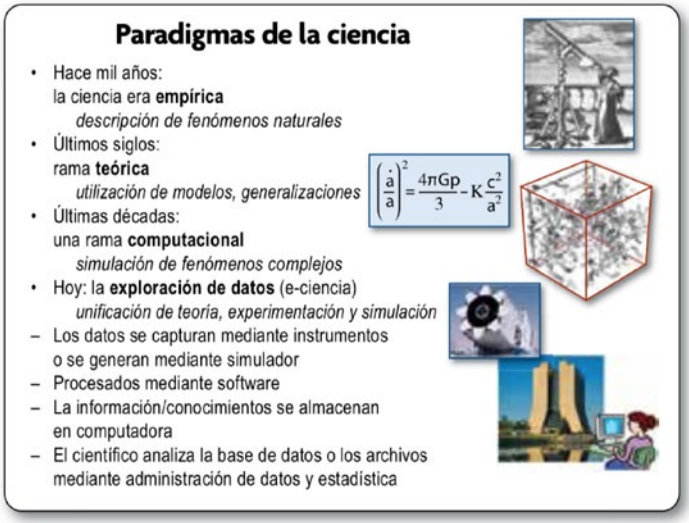

Ilustración 1. Paradigmas de la ciencia

En las etapas segunda y tercera, el ciclo de vida de los datos siguió un modelo que Briney (2015) resume en cuatro momentos: planeación del proyecto, recopilación de los datos, análisis y publicación (modelo que se ha venido manejando desde hace aproximadamente 400 años cuando se empezó con la publicación de artículos de investigación). En dicho ciclo, que concluía con la edición, generalmente impresa, no se consideraba a los datos como producto y solían perderse (Briney 2015, 299).

Como resultado del cuarto paradigma, el ciclo tradicional de vida de los datos se ve reformulado, sumándose al modelo los nuevos elementos que se muestran en la ilustración 2, donde no solamente se han agregado dos etapas más, sino que las anteriores se han fortalecido desde el momento en el que al mismo tiempo que se construye el proyecto de investigación se prepara también un plan de gestión de datos, la recopilación de los mismos y el análisis no se ven afectados,

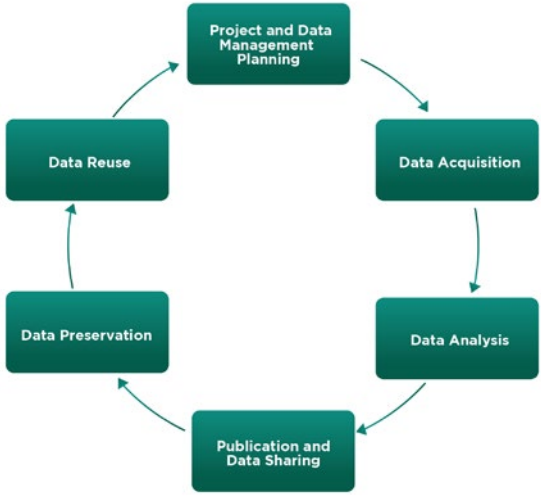

Ilustración 2. Nuevo ciclo de vida de los datos

pero al llegar a la etapa de publicación se agrega un proceso más: la compartición de los datos, puesto que, para publicar en revistas de alta calidad, los editores están requiriendo a los autores sus paquetes de datos con cada vez mayor frecuencia, no nada más el texto y las tablas o gráficas que acompañan al artículo, sino sus datos crudos para dar la opción de reproducir sus experimentos o retomarlos desde nuevas miradas en otras unidades de investigación; por ello también se agregan dos recuadros más con nuevas actividades, la preservación de los datos y la reutilización de los mismos.

Diversos estudios de la Universidad de Sheffield (Pinfield, Cox, y Smith 2014; Cox y Verbaan 2018; Cox *et al*. 2019) muestran que las grandes fuerzas que están impulsando el interés por la cultura de la gestión de datos de investigación (GDI) en la academia son:

> **Datos Masivos.** Este fenómeno fomenta nuevas formas de investigación puesto que "[..] crecen entre otras, la digitalización de las etapas de investigación, las capacidades de procesamiento, el trabajo colaborativo y a distancia entre investigadores, las herramientas de Internet para comunicar, recolectar y publicar, la experimentación con nuevas formas de visualización de datos" (Jonkowski, p. 7 2009, citado por Cox y Verbaan 2018, 42).
>
> **Editores de revistas.** Cada vez más las revistas científicas están estableciendo la entrega de los conjuntos de datos como requisito para la publicación de los artículos. Los lineamientos para su

inserción cuidadosa y su correspondiente documentación contemplan anexar, por ejemplo, los codebooks y diccionarios de datos (Cox y Verbaan 2018, 50); herramientas como grabaciones de video empleadas para complementar las descripciones tradicionales de los métodos experimentales, y de diversos tipos de visualizaciones de dos o tres dimensiones. [...] de tal forma que [...] el artículo se convierte así en una ventana no sólo para que un científico pueda comprender un resultado, sino para que lo reproduzca o lo amplíe. (Briney 2015, 195-96).

Seguridad y almacenamiento de datos. Con la producción de conocimiento basada en grandes cantidades de datos, estos requieren de atención escrupulosa, ya que están sujetos a pérdidas, contaminación digital y a dificultades en el acceso, por estas razones la protección de la seguridad, así como la necesidad de proporcionar facilidades de almacenamiento van siendo objeto de mayor atención en la academia.

Habilidades relacionadas con datos. El entorno de la investigación requiere cada vez más conocimiento de herramientas, plataformas, fuentes digitales y metodologías adaptadas a las prácticas actuales tanto en laboratorio individual como en el trabajo colaborativo con pares en diferentes latitudes. Los investigadores se están viendo en la necesidad de aprenderlas y aplicarlas en mayor medida.

Crisis de reproducibilidad. Cox y Verbaan (2018) indican que la agenda también se ha visto impactada por la crisis de confianza en torno a la integridad del trabajo de investigación y en consecuencia por la necesidad de mayor transparencia y acciones preventivas; al respecto ofrece ejemplos de las ciencias biomédicas y la psicología:

> En 2011 los científicos de un laboratorio farmacéutico escribieron un artículo científico revelando que les había resultado imposible reproducir los resultados de más de cuarenta y siete estudios sobre objetivos de fármacos que trataron de replicar […] al año siguiente se reportó en Nature que 47 de 53 estudios prominentes en investigación sobre cáncer no podían reproducirse tampoco […]. La conclusión que algunos comentaron fue que esos problemas no sólo eran resultado de un error causado por la complejidad de los experimentos, sino que se dieron a raíz de la presión de los investigadores para entregar resultados positivos para su publicación (Cox y Verbaan 2018, 43).

Políticas y programas de los gobiernos y agencias financiadoras. La publicación de la *Declaración sobre el Acceso a los Datos de Investigación* de la Organización para la Cooperación y Desarrollo Económico en 2004 y más adelante los *Principios y Lineamiento para el Acceso a los Datos de Investigación con Fondos Públicos* en 2007 favoreció una mayor conciencia en torno a la importancia de hacer los datos más accesibles

(OCDE 2007). A ello se han ido sumando otras obligaciones legales como la protección de datos y se han generado programas de apoyo a la creación de repositorios que antes no existían.

Movimiento de Acceso Abierto. La cultura de la apertura tiene que ver con todo un movimiento que incluye desde programas de cómputo, recursos de información, *MOCs* (Cursos masivos abiertos en línea) hasta academia abierta y e-Ciencia. Las metodologías están cambiando y por ello se requiere del manejo adecuado de los datos.

Cultura de los datos compartidos (*Data sharing*). El tema de la apertura y la evolución de la tecnología han dado lugar a dinámicas que antes se hubieran considerado impensables; es el caso del uso de datos generados en una investigación por parte de otros grupos académicos y científicos con la finalidad de reutilizarlos para trabajar nuevas hipótesis, observar otros patrones y generar nuevo conocimiento sin tener que duplicar esfuerzos innecesarios, y consiste en:

> […] la necesidad de compartir datos entre usuarios debidamente identificados y también proporcionar mecanismos y sistemas para permitir el acceso abierto a los datos donde sea apropiado. (Pinfield, Cox, y Smith 2014, 13)

La biblioteca como factor clave. En el entorno de la investigación intensiva en datos la contribución de las bibliotecas académicas es un motor fundamental, ya que puede y debe aportar experiencia

> en temas como elaboración de políticas, estructuración de proyectos basados en datos, confección de repositorios, acompañamiento a docentes, estudiantes e investigadores para la solución de sus necesidades de información en datos, capacitación, talleres y cursos de alfabetización en datos.

La pertinencia y los datos FAIR

De los datos masivos a los datos pertinentes

La producción de datos continúa creciendo y construyéndose; con lo cual la responsabilidad del profesional de la información se incrementa:

han asistido a seminarios, coloquios y congresos de la forma siguiente:

> […] todos estos bits y bytes [dan lugar a una] sobreproducción; entonces gira dialécticamente y se convierte en desinformación. Aquí [es donde] está el papel del bibliotecólogo: su nueva misión es ser un vigilante de este caos informativo, ser un vigilante de la estabilidad del conocimiento, de la información que ayuda a las sociedades a desarrollarse y no a desestabilizarse (Alfaro 2014).

Llevar estos saberes al mundo de los datos científicos representa participar de manera cercana con los investigadores, ofreciéndoles las soluciones más acordes a sus necesidades. Por ello, los siguientes pasos son integrarlos de manera coherente, logrando una masa crítica de conjuntos de datos en repositorios públicos y enlaces hacia y desde bases de datos científicas para mantener el registro académico en forma digital (Borgman 2008).

El brindar acceso sistematizado y pertinente a los datos, reporta, entre otros beneficios el transparentar los procesos de obtención y generación de los mismos en las instituciones; además:

> "[…] entre los investigadores promueve la colaboración entre grupos afines, y evita duplicaciones. Para la sociedad, los beneficios de hacer disponibles los datos mejoran la confianza en el sistema científico y supone un ejercicio de transparencia, rendición de cuentas y de responsabilidad en el uso de la inversión en I+D. Múltiples foros señalan el valor de los datos y la necesidad de la colaboración entre todos los agentes, ya que la ciencia conforma una infraestructura sociotécnica especialmente compleja, que implica tanto al sector público como al privado (Maredata 2019).

El concepto **FAIR**

¿Qué pasa con los datos que está produciendo cada investigador? ¿O con los datos que generan quienes que están elaborando sus tesis? ¿Qué pasa con esas entrevistas, con los cuestionarios, con la base de datos que desarrollaron, con las mediciones, los cruces de información, los diarios de investigación? Para que los datos no corran el riesgo de perderse, contaminarse o confundirse deberán, ante todo, tener cuatro características: ser localizables, accesibles, interoperables y reutilizables, por sus siglas en inglés: *Findable, Accessible, Interoperable and Reusable* (FAIR), tanto para los equipos informáticos como para las personas, lo cual tiene que ver con los datos pero también con las herramientas, vocabularios e infraestructuras que han de tomarse en cuenta para lograr adecuadamente los objetivos de descubrimiento y reutilización de información. (FORCE11 s. f.).

Para una mayor precisión de estos conceptos se ha recurrido a las descripciones que ofrecen Bezjak et al. (2018) y que se muestran en la tabla 1.

Tabla 1. Principios FAIR

PRINCIPIOS FAIR
Localizables: Para que los datos sean reutilizables, primero se deben localizar. Debería ser fácil encontrar los datos y los metadatos tanto para humanos como para las computadoras. El descubrimiento automático y fiable de conjuntos de datos y servicios depende de los identificadores persistentes (PIDs) legibles por máquinas y de sus metadatos
Accesibles: Los datos y metadatos deberían poder ser recuperados mediante su identificador utilizando un protocolo de comunicaciones abierto y estandarizado, que posiblemente incluya autenticación y autorización. Además, los metadatos deberían estar disponibles incluso si los datos ya no lo están.
Interoperables: Los datos deberían poder combinarse y usarse con otros datos o herramientas. Por lo tanto, el formato de los datos debe ser abierto e interpretable por distintas herramientas, incluyendo otros registros de datos. El concepto de interoperabilidad se aplica tanto en el nivel de los datos como en el de los metadatos. Por ejemplo, los (meta)datos deberían emplear un vocabulario que siga los principios FAIR.
Reutilizables: En última instancia, los principios FAIR buscan optimizar la reutilización de los datos. Para lograr esto, los metadatos y los datos deben estar correctamente descritos de modo que puedan ser replicados y/o combinados en diferentes entornos. Además, la reutilización de los (meta)datos debe estar indicada a través de licencias claras y accesibles.

Fuente: Elaboración original basada en Bezjak, Sonja, April Clyburne-Sherin, Philipp Conzett, Pedro Fernandes, Edit Görögh, Kerstin Helbig, Bianca Kramer, et al.(2019, 22)

Buenas prácticas y servicios de información

Señala Briney que una buena gestión comienza por los datos propios:

> En esencia GDI se trata de cuidar tus datos mejor de tal forma que no experimentes esas pequeñas frustraciones […] como tener problemas al localizar la documentación de un conjunto de datos en particular o problemas más grandes después de que se termina un proyecto, como la pérdida de datos (2015, 104).

Además, indica que un manejo eficiente de los datos de una investigación significa tener la posibilidad de:

> Encontrar un conjunto de datos en particular.
> Tener en orden todas las notas que necesitas.
> Poder prevenir una violación a la seguridad.
> Poder usar fácilmente los datos de un colaborador.
> Poder gestionar el caos de una cantidad de archivos digitales en permanente aumento.
> Ya sea que se trate de datos históricos de hace 50 años o de los datos críticos que recabaste para tu tesis doctoral la semana pasada (Briney 2015, 110).

De ahí que los investigadores podrían comenzar a desarrollar algunas buenas prácticas para la gestión de sus datos:

Compartición de datos. Se trata de todo lo relacionado con el cuidado de la propiedad intelectual, licenciamiento, contratos, patentes como protección previa a la liberación de los datos para evitar situaciones fuera de la legalidad con

editores, colegas, coautores y agencias que han financiado la investigación.

Preservación, almacenamiento y respaldo. Resguardo del original y copias de los archivos tanto en la propia computadora del investigador como en repositorios y otras opciones y equipos posibles. En esta etapa se definen también los formatos de los archivos, procurando que sean lo más abiertos posible. Tiene que ver con el respaldo de la información (hardware y software).

Reutilización. Implica las estrategias para localizar conjuntos de datos de otros investigadores y proyectos, y si una vez hecho esto los datos se pueden obtener de manera abierta o mediante un contrato, si es mediante colaboración con pares internos o de otras instituciones. En esta etapa del GDI la citación de datos es un tema fundamental, teniendo presente que "la reutilización de los datos es la última fase del ciclo de vida de los datos pero también el inicio de un ciclo nuevo." (Briney 2015, 3702)

Por otra parte, con respecto al rol específico del profesional de la información en estas acciones, Córdoba y Vega (2017) identifican una docena de características que deberían conformar el perfil de quien, desde la biblioteca, apoyaría la cultura de datos; esto es, un especialista con experiencia lo mismo en gestión técnica de datos para investigación y familiarizado con el manejo de las metodologías científicas que en tareas de uso, exploración, valoración y sistematización de

datos (capacidades que, cabe mencionar, se adquieren únicamente en el ejercicio cotidiano de la tarea investigadora). Sumado a estos saberes observan también la necesidad de habilidades tales como disposición para el trabajo en equipo, ética y consciencia.

En lo que se refiere a los servicios de datos que la biblioteca podría ofrecer se han identificado, de entre la amplia gama identificada en la literatura, los más significativos para integrarlos en las siguientes siete categorías:

Acceso pertinente a contenidos. Con base en sistemas de búsqueda y recomendación, estructurados a partir de perfiles de usuario para servicios de diseminación selectiva de información.

Referencia y búsqueda de datos. Comprende tanto el servicio de consulta tradicional como automatizado en fuentes locales, nacionales y mundiales.

Portales de conocimiento. Desarrollados mediante la aplicación de programas de minería de datos, curación y descubrimiento.

Orientación. Acompañamiento al académico en la selección de programas informáticos para datos estructurados, no estructurados y herramientas de visualización.

Analítica de usuarios. Investigación cuantitativa y cualitativa sobre los comportamientos de quienes asisten a la biblioteca.

Protección de datos. Se refiere a la orientación sobre *copyright*, normatividad. Por otra parte, también implica la asistencia para preservación de datos, así como desarrollo y atención a repositorios.

Alfabetización en datos. Denominación con la que se conoce a las acciones de capacitación en el uso de los datos y que involucra temas como "[…] mediciones (numéricas y no-numéricas), visualización, gráficos, gestión de procesos de investigación científica, *eResearch* y *eScience*" (Wanner 2017, 3); se ha considerado como "[…] una parte de la alfabetización informacional que permite: acceder, interpretar, valorar con sentido crítico, gestionar y usar éticamente los datos."(Calzada y Marzal 2013, 126)

Pensando en acciones complementarias, a futuro se podrían realizar estudios basados en datos para el análisis de redes sociales de la biblioteca; modelos predictivos (*machine learning*) para fortalecer las acciones a favor de los usuarios; clasificación con metadatos de los eventos organizados por la propia biblioteca aplicando procesamiento de palabras, indicadores estadísticos de frecuencia, longitud de palabras, posición de las palabras y otros diagnósticos de ese tipo.

Conclusiones

En el contexto posterior a la aparición de Internet, la producción de datos ha crecido a 1.8 *zetabytes* (1.8×10^{21} bytes)

de contenido digital en 2011, cantidad que se espera sea 50 veces mayor en 2020 (EMC 2011; Mearian 2011 citados por Briney 2015, 139).

En el ámbito académico se va fortaleciendo la conciencia en torno a la importancia que debe darse a la GDI. Con todo, aunque existe la necesidad entre los investigadores, aún no se hace del todo explícita y se va resolviendo localmente; si bien se ha generado la normatividad para regular este tipo de proceso, no se ha llevado a cabo como se esperaba y, finalmente, las bibliotecas pese a que tienen interés genuino en participar pareciera que no tienen la suficiente confianza en sus habilidades, capacidades y recursos (Cox *et al.* 2019, 10'51"), lo cual no es aceptable, puesto que los profesionales de la información cuentan con la preparación y la experiencia para integrarse en esta cultura y aportar al proceso de la investigación científica con las herramientas adecuadas para convertirse en custodios e intermediarios de datos.

La cultura de datos es un aprendizaje permanente que constituye un área de interés y de responsabilidad para los profesionales de la información. El experimentar buenas prácticas de datos que han sido probadas en otras bibliotecas alrededor del mundo fortalecerá la conciencia del rol del bibliotecario como actor importante en la academia, y serán fuente de motivación para nuevos servicios encaminados a apoyar al usuario científico, humanista o del ámbito de las artes, además de coadyuvar a la mayor calidad e integridad en los trabajos de investigación.

Bibliografía

Alfaro, H G. 2014. *Avatares de la información en la imagen. 11o. Seminario Hispano-mexicano de Investigación en Bibliotecología y Documentación*. Video. https://www.youtube.com/watch?v=dFDM2J7lcPc&index=5&list=PL1lcaDjIF8fiXbTZHWtP2VTojGsjnuSt.

Bezjak, Sonja, April Clyburne-Sherin, Philipp Conzett, Pedro Fernandes, Edit Görögh, Kerstin Helbig, Bianca Kramer, et al. 2019. *Manual de Capacitación Sobre Ciencia Abierta*. https://book.fosteropenscience.eu/es.

Borgman, Christine L. 2008. "Data, Disciplines, and Scholarly Publishing". *Learned Publishing* (21, 1): 29-38. https://doi.org/10.1087/095315108X254476.

Briney, Kristin. 2015. *Data Management for Researchers: Organize, maintain and share your data for research success*. Exeter, UK: Pelagic Publishing, Kindley Edition.

Calzada, Javier, y Miguel Ángel Marzal. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents". *Libri:* (63, 2). https://doi.org/10.1515/libri-2013-0010.

Córdoba, Mariana y Guadalupe Vega. 2017. "El perfil del bibliotecólogo responsable de la Gestión de Datos de Investigación". *Amontonamos las palabras: Blog de la Biblioteca de El Colegio de México* (blog). 2017. https://bdcv.hypotheses.org/558.

Cox, Andrew M., Mary Anne Kennan, Liz Lyon, Stephen Pienfield, y Laura Sbaffi. 2019. *Maturing research data services & the transformation of academic libraries.* Video. https://www.youtube.com/watch?v=yUTZ9t2j8-s.

Cox, Andrew, y Eddy Verbaan. 2018. *Exploring Research Data Management*. London: Facet Publishing.

FORCE11. 2019. "Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version B1.0". https://www.force11.org/fairprinciples.

Graham, Mark, y William Dutton. 2019. *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives*. 2.a ed. Oxford: Oxford University Press, USA.

Hey, Tony, Stewart Tansley, y Kristin Tolle, eds. 2014. *El cuarto paradigma. Descubrimiento científico intensivo en datos*. México: Universidad Autónoma Metropolitana.

International Telecommunication Unit, ed. 2018. *Measuring the Information Society Report*. Vol. 1. Geneva Switzerland: ITU.

Maredata. 2019. "Red Española sobre Datos de Investigación en Abierto". Maredata. https://maredata.net/index.php/presentacion/.

Martínez-Uribe, Luis, y Paz Fernández. 2015. "Servicios de datos: función estratégica de las bibliotecas del siglo XXI". *El Profesional de la Información* (24, 2): 193. https://doi.org/10.3145/epi.2015.mar.13.

Monash University. 2014. "Research Data Management Policy". http://policy.monash.edu.au/policy-bank/academic/research/research-data-management-policy.html.

OCDE, ed. 2007. "OECD Principles and Guidelines for Access to Research Data from Public Funding". *OCDE*. https://www.oecd.org/sti/inno/38500813.pdf.

Pinfield, Stephen, Andrew M. Cox, y Jen Smith. 2014. "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences". *PloS One* (9, 12). e114734. https://doi.org/10.1371/journal.pone.0114734.

Semeler, Alexandre Ribas, Adilson Luiz Pinto, y Helen Beatriz Frota Rozados. 2019. "Data Science in Data Librarianship: Core Competencies of a Data Librarian". *Journal of Librarianship and Information Science* (51, 3): 771-80. https://doi.org/10.1177/0961000617742465.

Wanner, Amanda. 2017. "Data Literacy Instruction in Academic Libraries: Best Practices for Librarians." *Archival and Information Studies Student Journal*, 1-17. https://doi.org/10.14288/sa.v1i1.186335.