

# Desiderata LAB

## ¿Qué es y qué importancia tiene la gestión de datos de investigación (GDI)?

*Julio Alonso-Arévalo*

*Biblioteca de la Facultad de Traducción y Documentación de la USAL*



***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” van a impulsar un crecimiento exponencial de los datos en el futuro más inmediato. La importancia de estos datos sobre patrones de comportamientos serán un recurso vital para el desarrollo social y económico de la sociedad, por lo que será necesaria la planificación, gestión y reutilización de estos datos.***

En el mundo se generan datos de manera continua, cada vez que enviamos un mensaje, hacemos una llamada o completamos una transacción, dejamos rastros digitales. Los “datos generados por el usuario” provienen de la participación de personas, por ejemplo, a través de plataformas de medios sociales, sistemas de localización

GPS o aplicaciones móviles entre otros. Sin duda, muchos.

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. 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 estos extremadamente grandes conjuntos de datos. De manera que casi todos los analistas consideran los “Big Data” como una de las tendencias de futuro que tendrán que tener en cuenta la mayoría de las empresas e instituciones. (Wanner, 2015)

Dentro de esta amplia variedad de datos están los datos abiertos, que hacen referencia a la información recopilada, elaborada o encargada por organismos públicos que puede utilizarse, modificarse y compartirse libremente por cualquier persona. Entre sus ventajas están el aumento en la transparencia y responsabilidad gubernamental, así como beneficios financieros tangibles para los ciudadanos, negocios y la sociedad civil en general. (Comisión Europea, 2018)

Fundamentalmente se trata de datos digitales de alta frecuencia, que se están convirtiendo en un elemento esencial para la competencia, y que 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. Procesarlos, entenderlos y transformarlos en decisiones de valor es el reto de los llamados «big data». Por darnos una idea, la importancia de los datos masivos es tal que el Foro Económico Mundial ha llegado a afirmar que “Los datos constituyen una nueva clase de activo económico, tan importante como lo fueron la moneda o el oro en los tiempos pasados.”

En el ámbito académico, el intercambio de datos tiene una importancia vital para la ciencia, ya que permite a los investigadores utilizar, analizar y reproducir los resultados de los demás para poder replicarlos o refutarlos con conocimiento de causa, ya que es posible tener acceso a los datos subyacentes que sustentan esa investigación. Además de tratarse de una cuestión puramente económica, ya que una de las tareas más costosas en un proceso de investigación es recopilar los datos necesarios para realizar el análisis. Incluso, muchas de estas compilaciones de datos ya han sido realizadas por investigaciones anteriores, pero no han sido compartidos públicamente, ni dispuestas en un repositorio para su reutilización por para otras investigaciones con diferentes propósitos, ya que quedan encerrados en repositorios de

datos, en las oficinas de los investigadores o en los organismos que lo crearon, lo que además perjudica a la capacidad de cita de los autores que además de ser citados por los documentos finales, lo pueden ser por los “sets de datos” que han proporcionado.

Pero los datos son más que números, fundamentalmente es información que puede ayudar a resolver problemas cotidianos de las personas. La promesa de un gobierno de datos abiertos, transparentes y accesibles contribuye a la mejora las capacidades y oportunidades para todos los ciudadanos, que puede cambiar el punto de vista sobre su entorno y empoderar a las personas para que puedan disponer de información útil para la toma de decisiones inteligentes basadas en datos. Por ejemplo, nos permitirían conocer la calidad del aire de nuestras ciudades, saber cuáles son las mejores zonas para hacer deporte, o bien conocer los flujos comerciales, el precio medio del alquiler en una zona y cuáles son los negocios de nuestra competencia, si estamos interesados en establecer nuestro propio negocio. Hablamos de las “Smart Cities” o ciudades inteligentes que son promesa de un futuro mejor basado en los datos.

De este modo, en un intento de abordar la cuestión, durante los últimos años, varios gobiernos, organizaciones científicas, tanto nacionales e internacionales han emitido declaraciones y políticas que subrayan la necesidad de un inmediato archivo de los datos. La gestión de los datos de investigación (GDI) se plantea como uno de los grandes retos que han de asumir durante los próximos años las bibliotecas científicas y de investigación. Pero cabe una duda. Esta práctica requiere de la transición desde la idea tradicional que se tiene de una “infraestructura de datos” centrada en torno a los resultados de la investigación a un sistema más robusto centrado en los datos de investigación. (Erway, R., L. Horton, et al., 2016). Hasta ahora las bibliotecas principalmente hemos gestionado los resultados finales de la investigación en forma de artículos o libros, y hoy gracias al esfuerzo realizado por bibliotecas y tecnólogos se calcula que la mitad de la investigación del mundo está libremente accesible a través de repositorios de acceso abierto. El siguiente desarrollo implicaría la gestión de todos los datos sub-



yacentes que se han utilizado en un proceso de investigación; esto incluye datos numéricos, estadísticos, cualitativos, presentaciones, etc. Incluso serían de enorme interés aquellos datos negativos de una investigación que no ha dado los resultados previstos. Todos estos datos son de enorme interés para la ciencia y su reproductividad.

Es tal su importancia, que las más importantes empresas tecnológicas y de edición del mundo se están posicionando, generando repositorios y servicios de datos. Así, compañías globales como Elsevier han diseñado depósitos de datos como DataSearch o Mendeley Data, ligada al famoso gestor de referencias; Clarivate, empresa que desarrolla Web of Knowledge (WoS), dispone de Data Citation Index, o Emerald que recientemente ha lanzado Emerald Open Research. Incluso Google ha presentado este año un buscador de conjuntos de datos para investigadores llamado Dataset Search, la herramienta -aun en beta – proporciona acceso a millones de conjuntos de datos en miles de repositorios de datos en la web de todo el mundo

¿Cómo gestionar los datos?

La Gestión de Datos de Investigación (GDI) -en inglés Research Data Management (RDM)- es un proceso diseñado para gestionar y difundir conjuntos de datos de alta calidad, que cumplan con los requisitos académicos, legales y éticos. Para ello es necesario crear aplicaciones, poner en marcha nuevos productos y ser-

vicios que mejoren la transparencia y la apertura interviniendo en todo el ciclo vital de los datos.

Cualquier investigación sólida debe ser reproducible y apoyarse sobre una base de datos sólidos y accesibles. Para que esto sea así, tanto en la práctica como en la teoría, se debe otorgar a los datos la debida importancia, y estar depositados en un registro académico duradero y sostenible que permita el acceso, uso, cita, reproducción y reutilización de los mismos. Por ello será necesario la intervención en todo el ciclo vital de los datos, incluso antes de que los propios datos se generen. Este ciclo comprende el procesamiento, análisis, preservación, acceso y reutilización de los datos. (Ver Fig. 1)

Lo primero sería desarrollar un plan de datos. Lo requiere conocer y documentar las siguientes cuestiones:

- ¿Quién posee los datos?
- ¿Qué requisitos son impuestos por otros?
- ¿Qué datos deben conservarse y por cuánto tiempo?
- ¿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?

Las bibliotecas estadounidenses ya están respondiendo a esta nueva misión y se calcula que aproximadamente la mitad de to-

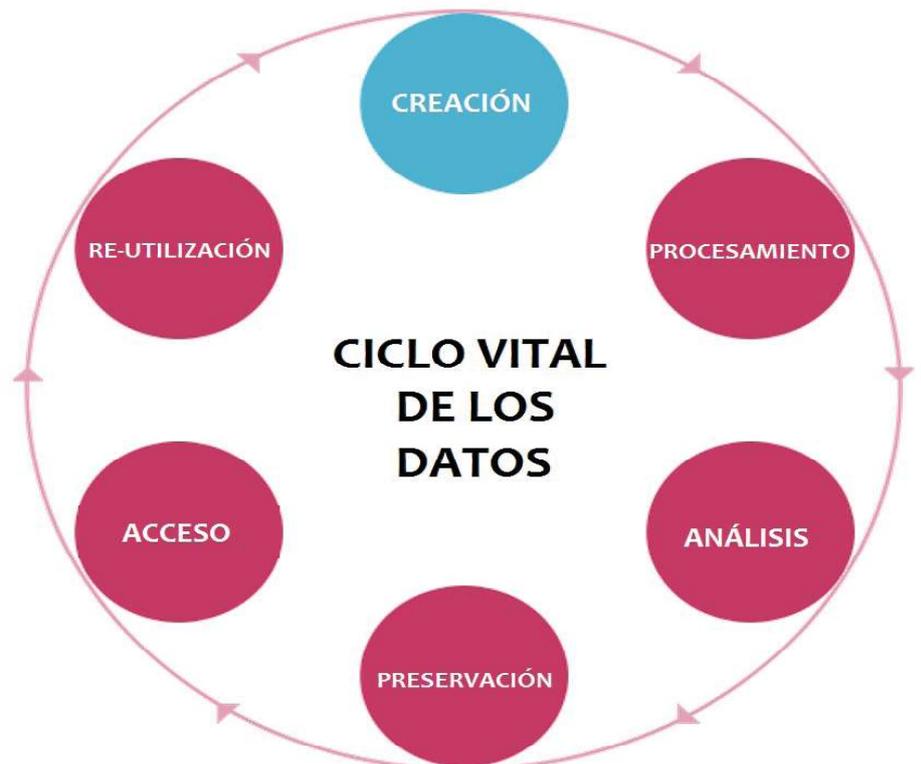


Fig. 1. Ciclo vital de los datos



das las bibliotecas universitarias tienen algún tipo de programa de apoyo a la gestión de datos. Para ayudar a iniciar un plan de datos de investigación LIBER, la coalición de bibliotecas de investigación europea, estableció 10 recomendaciones para que las bibliotecas empiecen a desarrollar un servicio de gestión de datos de investigación (GDI) (LIBER, 2012):

1. *Iniciar propuestas sobre Gestión de Datos de Investigación (RDM)*
2. *Participar en el desarrollo de estándares de metadatos*
3. *Difundir y desarrollar las habilidades necesarias para la gestión de datos entre el personal*
4. *Participar activamente en la investigación para el desarrollo de políticas*
5. *Mantener contactos y colaborar con los investigadores y grupos de investigación*
6. *Apoyar todo el ciclo de vida para los datos*
7. *Promover la investigación sobre datos mediante la aplicación de identificadores persistentes*
8. *Proporcionar un catálogo de datos institucional o repositorio*
9. *Participar en la práctica de gestión de datos en la disciplina de trabajo específica.*
10. *Ofertar o mediar en torno a cuestiones como el almacenamiento seguro*

Otras instituciones, asociaciones, y organizaciones como CLIR, JISC, Ithaka S+R y Sconul han publicado documentos y directrices para la adecuada gestión de datos de investigación donde podemos encontrar una buena ayuda para introducirnos a la Gestión de Datos de Investigación y en las mejores prácticas en la GDI. Algunos de estos documentos son:

- *Research Data Management : Principles, Practices, and Prospects. [e-Book]. New York, CLIR, 2013. <https://www.clir.org/pubs/reports/pub160/>*
- *Directions for Research Data Management in UK Universities. London, JISC, 2015. <https://www.fosteropenscience.eu/content/directions-research-data-management-uk-universities>*
- *Research Data Management Roles for Libraries. New York, Ithaka S+R, 2015 [https://pubs.](https://pubs.acs.org/doi/abs/10.1021/bk-2012-1110.ch008)*

*acs.org/doi/abs/10.1021/bk-2012-1110.ch008*

- *Research Data Management: Briefing for library directors, [e-Book] London: Sconul, 2015 <https://www.sconul.ac.uk/sites/default/files/documents/SCONUL%20RDM%20briefing.pdf>*
- *Turning FAIR Into Reality : Final Report/Action Plan From the Commission FAIR Data Expert Group. Brussels: European Commission Directorate General for Research and Innovation, 2018 [https://ec.europa.eu/info/sites/info/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf)*

### ¿Qué sectores profesionales participarían en la GDI?

La ciencia de datos está emergiendo como un campo que está revolucionando la ciencia y la industria por igual. El trabajo en casi todos los ámbitos se está orientando cada vez más hacia los datos, lo que afecta tanto a los empleos disponibles como a las competencias necesarias.

En cuanto a que gestores profesionales deberán realizar la GDI, la complejidad del proceso requerirá de la intervención de diferentes profesionales desde científicos de datos, hasta matemáticos, estadísticos, bibliotecarios y archiveros. La gestión de datos de investigación aparece como una de las tendencias clave en el trabajo que desarrollará el bibliotecario universitario del siglo XXI; si bien, al tratarse de un área emergente y en continua evolución, aún no están definidas las habilidades y competencias del bibliotecario de datos; en general la tarea del bibliotecario de datos implica a una amplia comunidad de bibliotecarios con diversas capacidades, antecedentes y responsabilidades profesionales. (Federer, 2018)

Los bibliotecarios llevamos 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. En palabras de MJ Tooey, bibliotecaria de la University of Maryland, “Tenemos el conjunto de habilidades necesarias para organizar las cosas. Entendemos los vocabularios controlados. En-





tendemos las ontologías. Sabemos cómo organizar la información. Hemos realizado la evolución de la palabra impresa y el encabezamiento de materia a los datos como un proceso natural”. Todo ello implicará acrecentar las competencias y habilidades de los bibliotecarios como uno de los segmentos profesionales más capacitados para la gestión de grandes cantidades de datos. (Taylor Stang, 2016).

## Conclusiones

*La gestión de datos de investigación es el nuevo gran reto para las bibliotecas universitarias y científicas que requerirá de la participación colaborativa de múltiples departamentos y organismos de la institución. Liderar el proceso de GDI supondrá contribución estratégica de primer orden para la valorización de la biblioteca en su institución. El proceso requerirá intervenir en todo el ciclo vital de los datos para identificar, etiquetar, gestionar y difundir conjuntos de datos de alta calidad, que cumplan con los requisitos académicos, legales y éticos. Además, los bibliotecarios estamos llamados a ser un elemento vital en liderar este proceso, ya que 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.*

*Sin embargo, 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 escuelas 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*

**<<Liderar el proceso de GDI supondrá contribución estratégica de primer orden para la valorización de la biblioteca en su institución>>**

## Bibliografía:

Comisión Europea. “Open Data Maturity in Europe Report 2018”. Comisión Europea, European Data Portal y Capgemini Invent, 2018 [https://www.europeandataportal.eu/sites/default/files/edp\\_landscaping\\_insight\\_report\\_n4\\_2018.pdf](https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.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/oclc-research-data-management-building-blocks-2016.pdf>

Federer, Lisa. “Defining Data Librarianship: A Survey of Competencies, Skills, and Training”. *Journal of Medical Librarianship (JMLA)*. Vol 106, No 3 (2018) <http://jmla.mlanet.org/ojs/jmla/article/view/306/655>

LIBER. *Ten recommendations for libraries to get started with research data management*, LIBER, 2012. <https://libereurope.eu/blog/2012/08/24/ten-recommendations-for-libraries-to-get-started-with-research-data-management/>

Stang, Taylor “Librarians: the new research data management experts : How growth in research data is spurring a shift in the librarian’s role” Elsevier, 2016 Disponible en: <https://www.elsevier.com/connect/librarians-the-new-research-data-management-experts>

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>

