



Palabra Clave (La Plata)
ISSN: 1853-9912
palabraclave@fahce.unlp.edu.ar
Universidad Nacional de La Plata
Argentina

La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas

Angelozzi, Silvina Marcela

La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas

Palabra Clave (La Plata), vol. 9, núm. 2, 2020

Universidad Nacional de La Plata, Argentina

Disponible en: <https://www.redalyc.org/articulo.oa?id=350562513004>

DOI: <https://doi.org/10.24215/18539912e091>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

La gestión de datos de investigación en abierto: introducción al rol emergente para las bibliotecas universitarias y científicas argentinas

Open research data management: introduction to the emerging roles for Argentine academic and research libraries

Silvina Marcela Angelozzi
Universidad Nacional de Córdoba. Facultad de Filosofía y
Humanidades. Escuela de Bibliotecología., Argentina
smangelozzi@gmail.com

DOI: <https://doi.org/10.24215/18539912e091>
Redalyc: <https://www.redalyc.org/articulo.oa?id=350562513004>

Recepción: 23 Julio 2019
Aprobación: 26 Diciembre 2019
Publicación: 10 Abril 2020

RESUMEN:

A partir de las políticas de acceso abierto adoptadas o promovidas por las agencias gubernamentales o públicas de financiamiento de las investigaciones, se establecieron normativas para el depósito de las publicaciones en repositorios abiertos. En algunos casos, ello va un paso más allá y se promueve también la apertura de los datos obtenidos durante el desarrollo de las investigaciones, para que puedan ser reutilizados por otros investigadores. En el presente artículo se presenta una introducción a la gestión de datos abiertos de investigación y se describen brevemente las tareas en las que pueden involucrarse las bibliotecas universitarias y científicas de Argentina a partir de la promulgación y reglamentación de la Ley 26.899.

PALABRAS CLAVE: Datos de investigación, Gestión de datos de investigación, Acceso abierto.

ABSTRACT:

The public agencies for research funding adopted regulations for the deposit of publications in open access repositories. In some cases, this goes further and also promote the opening of data obtained during the research process, so that they can be reused by other researchers. This article presents an introduction to the management of open research data and briefly describes the tasks in which the university and scientific libraries of Argentina can be involved, based on the promulgation and regulation of Law 26.899.

KEYWORDS: Research data, Research data management, Open access.

1. INTRODUCCIÓN

A partir de las políticas de acceso abierto adoptadas o promovidas por las agencias gubernamentales o públicas de financiamiento de las investigaciones, se establecieron normativas para el depósito de las publicaciones derivadas en repositorios abiertos. En algunos casos, ello va un paso más allá y se promueve también la apertura de los datos obtenidos durante el desarrollo de las investigaciones para que puedan ser reutilizados oportunamente. Ello hace que los investigadores deban ocuparse de mantener la integridad y comunicabilidad de los datos, no sólo para su propio uso sino también para que puedan ser puestos a disposición de terceros. Además, las instituciones donde desarrollan su labor deben proveer la infraestructura para gestionarlos apropiadamente. En este artículo, a partir de una revisión bibliográfica, se realiza una introducción a la gestión de datos de investigación y se describen brevemente las tareas que podrían asumir las bibliotecas y centros de información vinculados a instituciones de investigación en dos aspectos fundamentales: uno, el apoyo a los investigadores y otro, la participación en la infraestructura institucional de gestión de los datos.

2. ANTECEDENTES

Tradicionalmente, los investigadores han compartido sus datos de manera excepcional en dos circunstancias, como bien señala Sieber (1991): una, mediante los archivos académicos, gubernamentales, comerciales o de agencias públicas y otra, en los intercambios entre redes de investigadores que colaboran entre sí, conocidas comúnmente como “colegios invisibles”. Sin embargo, como sostiene esta autora, los archivos generalmente han sido vistos más desde una perspectiva histórica que desde una perspectiva del uso futuro de dichos datos.

A partir de las mayores facilidades de almacenamiento y recuperación de los datos en soporte digital, y luego a partir del movimiento de acceso abierto, el tema cobra un nuevo y mayor impulso. Así, por ejemplo, en el Reino Unido las agencias de financiamiento comenzaron a solicitar un plan formal de gestión de datos en los años noventa. En Estados Unidos, los National Institutes of Health (NIH) lo requieren desde 2003 para financiamientos superiores a los 500.000 dólares y la National Science Foundation (NSF) formaliza ese requisito para todos los proyectos a partir de 2010 y somete también el plan de datos a revisión de pares (Borgman, 2012, p.1060). La NSF ya desde 1985, solicitaba a los investigadores sociales poner a disposición los datos y la documentación relacionada en archivos públicos, un año después de finalizada la investigación (Sieber, 1991).

También, ciertas revistas científicas comenzaron a exigir la puesta a disposición de los datos de respaldo de los resultados de las investigaciones publicados en sus artículos, con mayor o menor grado de formalidad, por ejemplo, el depósito en un banco de datos *ad hoc* o un simple enlace a la página donde el autor los haya almacenado (Borgman, 2012). Es importante notar, como señala dicha autora, que la exigencia de la NSF se trata de “planes de *gestión* de datos” y no propiamente planes de “*data sharing*”, aunque sí se promueve a que los datos sean compartidos y se posibilite la reutilización (Borgman, 2012).

La Organización para la Cooperación y el Desarrollo Económicos en la reunión de París de 2004, mediante la *Declaration on Access to Research Data from Public Funding* (Organisation for Economic Co-operation and Development [OECD], 2004) se pronunció en favor del acceso abierto a los datos de investigación provenientes de investigaciones financiadas con fondos públicos y se comprometió a trabajar en un documento común de recomendaciones o normativas acordadas entre los diversos miembros, lo cual se concretó en el documento OECD *Principles and Guidelines for Access to Research Data from Public Funding* que se hizo público en 2007 (OECD, 2007). Esa serie de principios y recomendaciones persiguen el fin de facilitar el acceso eficiente y compartido a los conjuntos de datos provenientes de las actividades investigativas financiadas por los países miembros.

También en 2007, el Consejo de la Unión Europea remarcó la importancia de proveer acceso y permitir la posibilidad de re-uso de los datos de investigación e instó a los gobiernos de la Comunidad a coordinar los esfuerzos y las prácticas en torno a ello (Council of the European Union, 2007). En el programa de financiamiento de la investigación e innovación de la Comunidad Europea Horizonte 2020, se establecieron lineamientos para el acceso abierto tanto de publicaciones como de datos de investigación, estos últimos a través del *Open Research Data Pilot* (ORD) iniciado en 2017 (European Commission. Directorate General for Research & Innovation, 2017).

En el Reino Unido, la Research Information Network (2008) emitió un documento de recomendaciones en 2008, con la pretensión de constituir un marco amplio para promover las buenas prácticas de universidades, organismos de investigación, bibliotecas, editores, organismos de financiación e investigadores, bajo la premisa de que las ideas y el conocimiento producido con fondos públicos debe estar accesible para uso público, consulta, de la manera más amplia, rápida y eficiente posible.

En América Latina, la Comisión Económica para América Latina (CEPAL), a través de la Biblioteca Hernán Santa Cruz de Chile, formó parte del Proyecto *Leaders Activating Research Networks* (LEARN) impulsado por la League of European Research Universities (LERU). El objetivo fue el de construir una infraestructura de datos coordinada en Europa, Latinoamérica y el Caribe, tomando como base la Hoja de

Ruta LERU de Datos de Investigación. La propuesta consistió en el desarrollo de un modelo de gestión de datos de investigación y un kit de herramientas que facilitara su implementación. En el marco de este proyecto llevado a cabo durante 2016 y 2017 se realizó además un relevamiento de la situación en los países latinoamericanos y como resultado se obtuvo que entre las acciones llevadas a cabo en la región se destacaban: el establecimiento de leyes, el desarrollo de políticas en las agencias de financiamiento públicas, el desarrollo de repositorios de datos en universidades y en instituciones de investigación. Los mayores avances fueron detectados en Perú, Argentina, México, Brasil, Chile y Colombia (Andaur, 2016).

En Argentina, la Ley 26.899 de Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos promulgada en diciembre de 2013, establece que los proyectos financiados con fondos públicos que generen datos primarios deberán contener un plan de gestión de datos de modo de asegurar su accesibilidad a través de los repositorios destinados a su conservación, en los cuales deberán ser depositados en un plazo no mayor a los cinco años a partir de su recolección. Esta ley ha sido reglamentada a través de la Resolución 753 del entonces Ministerio de Ciencia, Tecnología e Innovación en noviembre de 2016. En dicha resolución se establecía un reglamento operativo que tiene por objetivo establecer los lineamientos básicos que deben cumplir los organismos e instituciones públicas que componen el Sistema Nacional de Ciencia Tecnología e Innovación (SNCTI). En el reglamento, se pautaron además los plazos de adecuación a la ley para las diferentes categorías de instituciones que conforman el sistema público de ciencia y tecnología del país.

Para las instituciones involucradas, uno de los principales desafíos consiste en generar y mantener archivos donde los datos de investigación puedan ser correctamente almacenados, identificados y preservados para que estén plenamente accesibles. Otro punto crucial es el involucramiento de los científicos y su cooperación (más allá de los mandatos obligatorios) en sistematizar sus planes de gestión de datos para luego ser volcados en los repositorios y en otorgar el consentimiento para compartirlos. Las bibliotecas en tanto preservadoras y difusoras de información científica han comenzado a implicarse gradualmente en ambos sentidos: en la concientización y capacitación de los investigadores y en la gestión de los archivos de datos.

3. ¿QUÉ SE ENTIENDE POR DATOS DE INVESTIGACIÓN?

Los NIH, una de las primeras instituciones en requerir un plan de gestión de datos de investigación en el mundo, los definen como: “material factual registrado, comúnmente aceptado en la comunidad científica como necesario para documentar, respaldar y validar los hallazgos de una investigación” (NIH, 2003). Se aclara también que se trata de datos finales, es decir no incluyen notas de laboratorio, conjuntos parciales de datos, análisis preliminares, borradores de artículos científicos, planes para futuras investigaciones, reportes de revisión de pares, comunicaciones con colegas u objetos físicos como especímenes de laboratorio. Tal como señalan Torres-Salinas, Robinson-García y Cabezas-Clavijo (2012) esta se ha convertido en una de las definiciones más aceptadas.

En el marco de la Ley 26.899 de nuestro país se entiende por dato de investigación o dato primario; “todo dato en bruto sobre los que se basa cualquier investigación y que puede o no ser publicado cuando se comunica un avance científico pero que son los que fundamentan un nuevo conocimiento” (Ley 26.899, Art.3).

Una de las primeras cuestiones que surge a partir de estas definiciones es la gran variedad de datos que se utilizan o se generan en las investigaciones, como pueden ser: mediciones tomadas por máquinas, una colección documental, un conjunto de entrevistas grabadas, un conjunto de respuestas a encuestas, registro de observaciones de campo, videograbaciones documentales, simulaciones computacionales, etc. Como puede apreciarse, existen diferencias asociadas a la naturaleza de los datos, a la forma de reproducibilidad y al nivel de procesamiento a los que estuvieron sometidos, lo cual tendrá influencia en las políticas de gestión que se establezcan (National Science Board, 2005).

En cuanto a su tipología existen algunas clasificaciones realizadas de acuerdo con diferentes criterios. Así, la NSF de Estados Unidos los clasifica a partir de su forma de obtención (National Science Board, 2005, p.19) en:

- Datos observacionales: obtenidos a partir de las observaciones realizadas en una investigación en un cierto momento, por lo tanto, no pueden volver a ser recolectados y su preservación es crítica.
- Datos computacionales: son resultado de ejecutar un modelo o simulación de computación, en este caso, puede no ser necesario preservar los outputs pero sí los parámetros del modelo (*hardware, software, inputs*).
- Datos experimentales: obtenidos usualmente a través de mediciones en laboratorios. Si bien las experiencias pueden ser reproducidas en muchos de los casos, su preservación es importante también en cuanto a los costos que puede suponer volver a colectarlos, o por la dificultad de conseguir exactamente las mismas condiciones.

En el mismo documento, se menciona otra clasificación relacionada con la etapa de procesamiento en que se encuentren:

- Datos en bruto: tal como son producidos o recolectados.
- Datos derivados: son aquellos construidos a partir del análisis, procesamiento o sistematización de datos en bruto, pudiendo generarse distintas versiones de datos.
- Datos intermedios: colectados en etapas preliminares de la investigación.
- Datos finales: el conjunto de datos seleccionados que finalmente se utilizan para validar los resultados.

En el reporte de Swan y Brown (2008) realizado para la red RIN del Reino Unido, se agrega el criterio de clasificación según el objetivo de recolección:

- Propósito específico: para un proyecto particular y pueden tener poco de interés fuera de dicho proyecto.
- De mediano alcance: pueden proveer información de interés o ser de referencia para una disciplina o grupo de disciplinas.
- De amplio interés: pueden constituir una referencia canónica para un gran número de investigadores en el mundo.

Torres-Salinas, Robinson-García y Cabezas-Clavijo (2012, p.175) mencionan también una clasificación de acuerdo con su formato: texto, imagen, audio y/o video.

Estas clasificaciones si bien brindan un panorama, siempre son parciales y no pueden dar cuenta de la diversidad de datos que se producen en el contexto de las investigaciones. La variedad y complejidad hacen que establecer las políticas de conservación y gestión de datos sea dificultoso y que no todos los modelos propuestos puedan ser extrapolados de una disciplina a otra. Como señala Borgman (2012, p.1072), los productores y los interesados en compartir datos no pueden ponerse totalmente de acuerdo aún en *qué* son exactamente los datos, ya que pueden ser mucho más complejos que un conjunto de números en una página.

4. ¿QUÉ IMPLICA LA GESTIÓN DE LOS DATOS DE INVESTIGACIÓN?

Para que los datos de investigación estén en acceso abierto y exista la posibilidad real de ser localizados y reutilizados por otros investigadores, se requiere de la generación de una infraestructura de gestión.

En un estudio llevado a cabo por OCLC, donde se analizaron las prácticas de gestión de datos de universidades norteamericanas, se identificaron tres categorías principales de servicios en torno a ellas: educación, experticia y curaduría (Bryant, Lavoie, & Malpas, 2017).

El *servicio de educación* pretende concientizar a los investigadores y a los demás actores involucrados sobre la importancia de preservar y compartir los datos, de respetar las normativas de la institución o de la agencia de financiamiento, de aplicar los protocolos. Asimismo, se les ofrece capacitación específica para diseñar planes de manejo responsable de los datos que producen.

El *servicio de experticia* tiene a su cargo ofrecer soporte y soluciones ante los problemas que puedan surgir en la práctica. Para ello pueden usarse variados canales de comunicación entre los expertos y los investigadores, privilegiando el contacto persona a persona.

El *servicio de curaduría* provee la infraestructura tecnológica y los servicios relacionados durante todo el ciclo de investigación. Incluye el tratamiento de los datos cuando son generados y el soporte posterior para asegurar la permanencia a largo plazo luego de concluida la investigación. Implica una serie de operaciones como almacenamiento seguro, identificadores persistentes, asignación de metadatos, identificación de las sucesivas versiones y preservación a largo plazo. La curaduría asegura la consistencia, completitud, persistencia y acceso a los datos.

Para que el modelo basado en la tríada educación, experticia y curaduría funcione, es importante que las instituciones cuenten con una política explícita y clara en cuanto a qué, cómo, para qué conservar y compartir los datos. El modo de organizar y llevar a cabo los tres servicios no responde a un esquema universal o rígido, sino que estará fuertemente asociado a la política particular de cada institución y variará según cada contexto.

En el marco del Proyecto LEARN (2017) se elaboró un documento modelo de políticas de gestión de datos donde los ítems a considerar son:

- Jurisdicción de aplicación
- Gestión de derechos de propiedad intelectual: de los investigadores, institución de filiación, agencias de financiamiento.
- Cuestiones operativas relativas al archivo de los datos y gestión de los repositorios (integridad, identificación, trazabilidad, accesibilidad, interoperabilidad, tiempo de resguardo, normativas de descripción y citación)
- Responsabilidades, tareas y derechos de los investigadores: respeto por las políticas institucionales, elaboración del plan de gestión de datos, posibilitar la reutilización de los datos.
- Responsabilidades, tareas y derechos de la institución: entre otras proveer capacitación e infraestructura para archivo y acceso a los datos, soporte permanente.

5. EL PLAN DE GESTIÓN DE DATOS

Las políticas de gestión de datos se redactan en términos generales y su concreción se lleva a cabo a través de los comúnmente denominados “Planes de gestión de datos” (PGD).

Las agencias de financiamiento exigen a los investigadores que reciben los subsidios que elaboren e informen el plan de gestión de datos para cada proyecto financiado. Estos investigadores pueden pertenecer a diferentes disciplinas e instituciones, por lo cual el plan deberá contemplar las políticas de la propia institución además de la del agente financiador. Por lo tanto, es deseable que las políticas institucionales contemplen esta situación de terceras partes involucradas y sean lo suficientemente flexibles.

Por ejemplo, en las *Directrices para la Gestión de Datos en Horizonte 2020* (2014, pp. 4-5) se señala:

Un PGD describe el ciclo de vida de la gestión de todos aquellos conjuntos de datos que sean recopilados, procesados o generados por el proyecto de investigación. Es un documento en el que se describe cómo serán manipulados los datos en el transcurso del proyecto de investigación e incluso tras su conclusión, describiendo qué datos serán recopilados, procesados o generados, qué metodología y estándares se utilizarán, cómo se compartirán y serán abiertos, y por último cómo se conservarán y preservarán. El PGD no es un documento definitivo, sino que irá evolucionando y adquiriendo más precisión y entidad durante el período de vigencia del proyecto. (p.4-5)

La NSF (2018) en el documento guía para la solicitud de subsidios establece que debe adjuntarse un documento de no más de dos páginas donde conste un PGD, que contemple los siguientes aspectos: tipos de datos que se producirán; estándares que se utilizarán para el registro y los metadatos; políticas de acceso, compartición y protección; políticas para reutilización, distribución y generación de datos derivados; detalles sobre el archivo y preservación. El PGD es evaluado en conjunto con la propuesta de investigación.

En el Reino Unido, el Digital Curation Centre (2018) facilita un documento propio con indicaciones para completar un PGD donde los ítems a incluir son básicamente coincidentes con los de la NSF expuestos en el párrafo anterior. Este centro facilita además una herramienta online para la creación del PGD teniendo en cuenta los requisitos y políticas de varias instituciones y universidades.

En Argentina, el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, 2017) exige desde el 29 de noviembre de 2017 que los equipos de investigación que se desempeñan en su ámbito deben presentar un plan para la gestión de los datos que describa el tratamiento que van a recibir los datos de investigación recolectados o producidos en el curso de un proyecto de investigación a ser financiado. La confección del PGD se realiza desde la plataforma virtual provista por el organismo y a la que se accede a través de la intranet del CONICET (2017) donde se provee de una guía para la elaboración. Los aspectos contemplados son:

- Recolección y creación de datos: corresponde a la descripción de los datos que serán generados o recolectados.
- Procedencia y Metadatos: documenta la procedencia o información relacionada con entidades, actividades o personas involucradas en la producción de los datos y que puede servir para determinar su calidad o confiabilidad. Se detalla el uso de esquemas o estándares de metadatos.
- Organización, almacenamiento y resguardo: se releva información acerca del volumen de datos que se prevé generar durante la investigación, cómo se almacenarán y de qué forma se realizarán las copias de seguridad.
- Acceso, intercambio, reutilización: se registra la información relativa a la forma de acceso e intercambio de datos y si se aplican restricciones.
- Ética y conformidad legal: donde se describe acerca de las necesidades de contar con el consentimiento informado, asegurar anonimato y otros aspectos relativos a la confidencialidad.
- Responsabilidades y recursos: donde se asignan tareas y responsables dentro del grupo de investigación necesarios para la implementación del Plan de Gestión de Datos.

6. REPOSITARIOS DE DATOS DE INVESTIGACIÓN

Los datos de investigación, convenientemente descriptos y con metadatos asignados para su posterior recuperación, se depositan en los denominados repositorios de datos. Kowalczyk & Shankar (2011, p. 261) definen un repositorio como “un sistema y un conjunto de servicios diseñados como archivo de datos digitales con contexto, fijeza y persistencia”. Estos autores distinguen cuatro modelos de arquitectura: *centralizados*, que dependen de una sola institución; *descentralizados*, que dependen de varias instituciones, pero se centraliza el archivo físico de los datos; *federados*, entre varias instituciones con archivo físico distribuido entre las instituciones participantes con una única interfaz de recuperación. Otro modelo más complejo es el de *ciberestructura*, basado en las capacidades de Internet, con los datos distribuidos en la red y permanentemente accesibles.

Los repositorios, además de asegurar la preservación a largo plazo, deben proveer acceso a los conjuntos de datos. Para ello, deben ser identificados y descriptos con metadatos “ricos” que aporten el contexto para su posterior interpretación y deben proveer interfaces amigables para la recuperación. Es altamente deseable

que se puedan establecer conexiones con las investigaciones que les dieron origen y con las publicaciones que derivaron de esos datos, por lo cual la interoperabilidad es de vital importancia.

7. ACTORES, ROLES Y RESPONSABILIDADES

En el ciclo de vida de los datos, desde su origen hasta su posible destrucción en caso de que así se decida, intervienen distintos actores con roles y responsabilidades definidos.

En el Proyecto LEARN (2017) se identifican como actores principales los investigadores, las instituciones de filiación de los investigadores, los agentes de financiamiento y las entidades de soporte a la investigación como bibliotecas, servicios tecnológicos, centros de apoyo a la investigación. De acuerdo con el proyecto mencionado, los roles y responsabilidades pueden resumirse de la siguiente forma:

Investigadores:

- generan los datos (crean, recolectan, recogen)
- diseñan el PGD de acuerdo con las políticas de las instituciones y agentes de financiamiento;
- deciden grado de apertura, confidencialidad, restricciones de acuerdo con el tipo de estudio;
- describen y archivan los datos en los repositorios;
- reutilizan datos generados por otros investigadores.

Instituciones de investigación:

- establecen las políticas generales y flexibles de apertura y compartición de datos;
- armonizan políticas con entes financiadores e intereses de los investigadores;
- proveen infraestructura tecnológica y administrativa de archivo y gestión de datos (archivo, preservación, acceso, expurgo, estándares);
- proveen apoyo y capacitación a los investigadores;
- asumen todo o parte de los costos de archivo, preservación y acceso;
- promueven la cultura de datos abiertos.

Agencias de financiamiento:

- establecen las políticas de apertura y compartición de datos;
- proveen modelos de PGD para ser empleados por los investigadores;
- pueden proveer repositorios;
- promueven la cultura de datos abiertos.

Centros de apoyo a la investigación:

- apoyo a las instituciones e investigadores en la curaduría de datos (descripción, metadatos);
- alfabetización en datos;
- proveen servicios de búsqueda de datos;
- promueven la reutilización de los datos;
- promueven la cultura de datos abiertos.

Briney, Golen & Zilinski (2017) incluyen también a las revistas y publicaciones científicas como un actor más, ya que muchas exigen el depósito de los datos de investigación como apoyatura de lo sostenido en los artículos que publican.

Lyons (2007) detalla además los derechos de los involucrados: los investigadores tienen derecho a ser los primeros usuarios de los datos, a ser reconocidos y citados, a recibir entrenamiento y capacitación en gestión

de datos; las instituciones que albergan a los investigadores tienen derecho a recibir una copia de los datos; los entes financiadores tienen derecho a implementar políticas de datos y a requerir que los investigadores las cumplan; los editores tienen derecho a esperar que los datos sean confiables como soporte de la publicación y a requerir el depósito en repositorios a largo plazo. Los usuarios posteriores tienen derecho a tener una licencia de uso y a acceder a metadatos de calidad que favorezcan la interpretación.

8. ROL EMERGENTE PARA LAS BIBLIOTECAS

Si se piensa tanto desde el modelo propuesto por Bryant, Lavoie, & Malpas (2017) con las tres dimensiones: educación, experticia y curaduría, como desde los roles identificados por el Proyecto LEARN (2017) para los centros de apoyo, surgen múltiples posibilidades de acción para las bibliotecas de investigación. Ello entraña necesidades específicas de capacitación del personal de las bibliotecas para el desempeño de las diferentes tareas.

Uno de los aspectos cruciales, es la capacitación de los investigadores o *alfabetización en datos*. Las bibliotecas tienen vasta experiencia en alfabetización en información, con numerosas formas y estrategias tanto presenciales como virtuales: confección de guías, tutoriales, encuentros presenciales, referencia por chat, etc. Así, pueden capacitar a los investigadores en:

- confección de los planes de gestión de datos exigidas por entes financiadores;
- adición de licencias de uso para proteger y compartir los datos;
- gestión de archivos personales de datos (identificación, copias de respaldo, control de versiones);
- creación de información contextual para los datos;
- aspectos éticos de la reutilización de datos;
- normativas de citación de conjuntos de datos.

Para ello, el personal de la biblioteca deberá conocer y mantenerse actualizado en cuanto a las normativas emanadas de los entes financiadores. Estos entes generalmente ponen a disposición los textos de las normativas, así como tutoriales en línea y ofrecen también capacitaciones presenciales. En este sentido, en la Argentina, CONICET creó la Red Federal de Especialistas que tiene más de 50 miembros, la mayoría de ellos bibliotecarios y profesionales de la información distribuidos en las distintas dependencias del CONICET, que se ocupan de procesar la producción publicada por los investigadores de las unidades a las que pertenecen y que cuentan con numerosas instancias de encuentro y capacitación permanentes. Dicha red, ya constituida puede ser extendida para la gestión de datos de investigación. Se han realizado talleres en cuanto a ciencia abierta y datos de investigación en distintas provincias, principalmente para investigadores CONICET y autoridades, pero que han estado abiertos también para bibliotecarios y profesionales de la información.

Además, se debe estar al corriente de las licencias abiertas disponibles para conjuntos de datos y sobre los aspectos éticos de su reutilización con su correspondiente reconocimiento y citación.

En cuanto a la citación de conjuntos de datos de investigación, se abre para los bibliotecarios una vasta oportunidad para el análisis y la discusión, ya que los estilos de citaciones más difundidos proveen ejemplos muy básicos que difícilmente puedan abarcar la diversidad y complejidad que presentan.

En las cuestiones más operativas de la gestión del archivo y preservación de datos, las bibliotecas pueden asumir el liderazgo o proveer apoyo a las instituciones en la curaduría de datos en los siguientes aspectos:

- descripción y documentación de los datos;
- elección de esquemas de metadatos a adoptar;
- normalización en la publicación de datos;
- desarrollo de ontologías;
- digitalización de sets de datos históricos que se posean en otros formatos;

- integración del repositorio institucional y catálogo de biblioteca con los conjuntos de datos cuando corresponda.

Todo ello, requiere también la capacitación de los bibliotecarios en esquemas de metadatos apropiados para este tipo de información, ya que es de naturaleza diferente a la de carácter bibliográfico con la que están más familiarizados. Los datos necesitan ser descriptos de modo de lograr una correcta identificación, conservar su integridad, asegurar la trazabilidad de las diferentes versiones y lograr la máxima accesibilidad e interoperabilidad. También los sistemas bibliotecarios deberán ajustarse de modo que permitan enlazar la información bibliográfica publicada a partir de las investigaciones con los datos que les dan sustento, de modo de potenciar su utilidad.

Todo ello, requiere también la capacitación de los bibliotecarios en esquemas de metadatos apropiados para este tipo de información, ya que es de naturaleza diferente a la de carácter bibliográfico con la que están más familiarizados. Los datos necesitan ser descriptos de modo de lograr una correcta identificación, conservar su integridad, asegurar la trazabilidad de las diferentes versiones y lograr la máxima accesibilidad e interoperabilidad. También los sistemas bibliotecarios deberán ajustarse de modo que permitan enlazar la información bibliográfica publicada a partir de las investigaciones con los datos que les dan sustento, de modo de potenciar su utilidad.

Entre los servicios a partir de los datos que pueden generarse se cuentan:

- búsqueda y recuperación de conjuntos de datos;
- localización de buscadores y metabuscadores de datos;
- localización de repositorios existentes, sus formas de acceso y uso de los datos que sean de relevancia para la comunidad de investigadores;
- difusión de los datos para incentivar reutilización;
- métricas del uso de los datos.

Para ello, los bibliotecarios deben estar familiarizados con los directorios de repositorios como por ejemplo Re3Data.Org, Odisca y OpenDoar; con buscadores como los que proponen OpenAIRE, Mendeley Data, DataSearch de Elsevier (versión Beta) y Google Dataset Search (versión Beta) y productos comerciales como Data Citation Index (DCI) de Clarivate Analytics y Scopus. Estas fuentes pueden ser utilizadas también para realizar redifusión de los datos y para hacer el seguimiento de citas e impacto.

En relación con las políticas institucionales, sería importante que la biblioteca sea convocada a participar en las discusiones del establecimiento de lineamientos en los diversos aspectos que se han mencionado en los párrafos anteriores. Como contraparte, es importante que los profesionales que se desempeñan en las bibliotecas demuestren las competencias adecuadas y la disposición para involucrarse en un campo que presenta nuevos desafíos. En este sentido, hay que reconocer que los bibliotecarios no siempre son percibidos por la comunidad científica como potenciales colaboradores cercanos.

A todo lo anterior, se puede agregar la intervención activa que puede tener la biblioteca de investigación en la promoción de la cultura de datos abiertos, para generar conciencia de su importancia en los investigadores, de modo de promover una efectiva reutilización de los datos para la producción de nuevo conocimiento y favorecer la transición hacia un modelo de ciencia abierta.

A MODO DE CIERRE

La participación en la gestión de datos de investigación constituye un reto a la vez que una oportunidad para las bibliotecas universitarias y de investigación, para ello es necesario que su personal cuente con las competencias y los conocimientos necesarios para asumir estos roles. Si bien muchas de las actividades se corresponden con lo que ya saben hacer, es importante ahondar en la naturaleza de los datos primarios y

en las diferencias que guardan en forma y escala con las publicaciones con las que están acostumbrados a trabajar. Se abre así, una posibilidad de especialización para los profesionales bibliotecarios. También se debe tener presente que es necesario trabajar en equipos interdisciplinarios y en conjunto con otras dependencias u organizaciones no bibliotecarias.

En cuanto a la participación efectiva de las bibliotecas en la gestión de datos existe ya experiencia en bibliotecas norteamericanas y europeas como lo evidencian los estudios de Tenopir, Birch & Allard (2012); Tenopir et al (2017); Cox & Pinfield (2014). Las bibliotecas argentinas, atendiendo a sus contextos particulares, tienen la oportunidad de iniciar un camino de cooperación con los investigadores y con las instituciones que las albergan en este desafío emergente en el país.

REFERENCIAS

- Andaur, G. (2016). *Panorama de la gestión de datos de investigación en América Latina y El Caribe*. Recuperado de <http://learn-rdm.eu/es/gestion-de-datos-de-investigacion-en-america-latina/>
- Borgman, C. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science*, 63(6), 1059–1078. doi:10.1002/asi.22634
- Briney, K.; Goben, A. & Zilinski, L. (2017). Institutional, Funder and Journal Data Policies. En L. Johnston, *Curating research data. Volumen I* (pp.61-78). Chicago: ACRL.
- Bryant, R.; Lavoie, B. & Malpas, C. (2017). *A tour of the research data management (RDM) service space. The realities of research data management, Part 1*. Dublin, Ohio: OCLC. doi: <https://10.25333/C3T92S>.
- CONICET. (2017). *Confeción de un plan de gestión de datos. PGD CONICET. Guía Usuario*. Buenos Aires: CONICET. Recuperado de <http://pgd.conicet.gov.ar/files/guia-usuario-pgd-conicet.pdf>
- Council of the European Union. (2007). *Council conclusions on scientific information in the digital age: access, dissemination and preservation*. Bruselas: CEU. Recuperado de <http://data.consilium.europa.eu/doc/document/ST-15362-2007-INIT/en/pdf>
- Cox, A. & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of librarianship and information science*, 46(4), 299-316. doi: <https://10.1177/0961000613492542>
- Digital Curation Centre. (2018). *Digital Curation Centre: DCC Template*. Recuperado de https://dmponline.dcc.ac.uk/template_export/1638514350.pdf
- Directrices para la Gestión de Datos en Horizonte 2020*. (2014). Recuperado de https://www.consorcioadrono.es/wp-content/uploads/2017/05/directrices_gestion_datos_horizon_2020_es.pdf
- European Commission. Directorate General for Research & Innovation. (2017). *H2020 Programme. Guidelines to de Rules on Open Access to Scientific Publications and Research Data in Horizon 2020*. Recuperado de https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Kowalczyk, S. & Shankar, K. (2011). Data sharing in the sciences. *Annual review of information science and technology*, 45 (1), 247-294. Recuperado de <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.2011.1440450113>
- LEARN Toolkit of Best Practice for Research Data Management*. (2017). doi: <https://10.14324/000.learn.00>
- Lyon, L. (2007). *Dealing with Data: roles, rights, responsibilities and relationships. Consultancy report*. . UKOLN. Recuperado de https://purehost.bath.ac.uk/ws/portalfiles/portal/419529/dealing_with_data_report-final.pdf
- National Institutes of Health. (2003). *NIH data sharing policy and implementation guidance*. Recuperado de https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#goals
- National Science Board. (2005). *Long-lived digital data collection: enabling research and education in the 21st century*. Arlington: NSF. Recuperado de <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>
- National Science Foundation. (2018). *Proposal & award policies and procedures guide*. Alexandria, Virginia: NSF. Recuperado de https://www.nsf.gov/pubs/policydocs/pappg18_1/nsf18_1.pdf

- OECD. (2004). *Science, technology and innovation for the 21st century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level*. Paris: OECD. Recuperado de <https://www.oecd.org/sti/sciencetechnologyandinnovationforthe21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm>
- OECD. (2007). *OECD Principles and guidelines for access to research data from public funding*. Paris: OECD. Recuperado de <https://www.oecd.org/sti/sci-tech/38500813.pdf>
- Research Information Network (UK). (2008). *Stewardship of digital research data: a framework of principles and guidelines*. London: RIN. Recuperado de <http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf>
- Sieber, J. (1991). *Sharing social science data: advantages and challenges*. Newbury Park: Sage.
- Swan, A. & Brown, S. (2008). *To share or not to share: publication and quality assurance of research data outputs. A report commissioned by the Research Information Network*. London: RIN. Recuperado de https://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf
- Tenopir, C.; Birch, B. & Allard, S. (2012). *Academic libraries and research data services: Current practices and plans for the future*. Chicago: ACRL. Recuperado de http://www.ala.org/acrl/sites/ala.org/acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Tenopir, C. *et al.* (2017) Research data services in European academic research libraries. *LIBER quarterly*, 27(1), 23–44. doi: <https://10.18352/lq.10180>
- Torres-Salinas, D.; Robinson García, N. y Cabezas Clavijo, A. (2012). Compartir los datos de investigación: introducción al data sharing. *El profesional de la información*, 21(2), 173-184. doi: <https://10.3145/epi.2012.mar.08>