

Universidad Nacional de Rosario
Facultad de Humanidades y Artes
Licenciatura en Bibliotecología y Documentación

Datos abiertos de investigación y bibliotecas científicas:

nociones, relaciones, actividades y competencias.

Alumno: Bibl. Fernando Gabriel Díaz Pacífico
Legajo N° D-0324/7
Año de ingreso: 2015

Tutor: Profesor Lic. Nicolás Tripaldi

Año 2020

Dedico este trabajo a:

Mi madre, Gladys, por apoyarme y brindarme el ejemplo de la constancia y el valor del trabajo.

A mis Profesores Nicolás y Ana, por edificar, por orientar, por sostener, por promover...

A Alcira, que es Maestra de tantas cosas.

A Rita, Jonathan, Micaela, Paulina y Silvia, por acompañar.

Resumen

Los datos de investigación son cada vez más importantes en el marco de la ciencia abierta. La gestión de estos activos de información representa un gran desafío para las instituciones de investigación y una oportunidad para las bibliotecas. Este trabajo analiza los principales elementos de la gestión de datos de investigación (GDI), y establece la relación de las bibliotecas con este proceso, identificando elementos de la GDI, actividades y competencias de los bibliotecarios de datos. Se revisa la literatura publicada entre 2012-2020 para identificar las variables de interés. El análisis se complementa con publicaciones oficiales. La inclusión de la legislación argentina relativa al tema sitúa la investigación. En base a la información relevada, surge que las bibliotecas anglosajonas y europeas son pioneras y están desarrollando actividades y prestando servicios especializados, dándole forma al rol del bibliotecario de datos. La baja incidencia de estudios parece indicar que las bibliotecas de América Latina todavía no alcanzaron ese nivel de desarrollo. En Argentina la legislación constituye un marco propicio para que las bibliotecas asuman este desafío. La colaboración con distintas áreas, la existencia de políticas institucionales, la formación académica, la actualización continua y la disponibilidad de recursos, surgen como factores trascendentes para las posibilidades de las bibliotecas. El análisis identifica un cuerpo de conocimientos, competencias y habilidades relacionados con la GDI desde la perspectiva bibliotecaria.

Palabras clave: datos de investigación; bibliotecario de datos; gestión de datos de investigación; bibliotecas académicas; conocimientos, competencias y habilidades; servicios de datos de investigación.

Abstract

Research data is increasingly important in the open science framework. Managing these information assets represents a great challenge for research institutions and an opportunity for libraries. This work analyzes the main elements of research data management (RDM), and establishes the relationship of libraries with this process, identifying elements of RDM, activities and competencies of data librarians. The literature published between 2012-2020 is reviewed in order to identify the variables of interest. The analysis is complemented by official publications. The inclusion of Argentine legislation gives context to research. Based on the information gathered, it appears that North America and European libraries are pioneers developing activities and providing specialized services, shaping the role of the data librarian. The low incidence of studies seems to indicate that libraries in Latin America have not yet reached that level of development. In Argentina, the legislation constitutes a favorable framework for libraries to take on this challenge. Collaboration with different areas, existence of institutional policies, academic training, continuous training and the availability of resources, emerge as important factors for the possibilities of libraries. The analysis identifies a body of knowledge, competencies, and skills related to RDM from a librarianship perspective.

Keywords: research data; data librarians; research data management; academic libraries; knowledge and skills; research data services.

TABLA DE CONTENIDO

- 1. Prólogo**
- 2. Introducción**
- 3. La apertura de la ciencia**
- 4. Beneficios de lo abierto**
- 5. Los datos, las bibliotecas y los bibliotecarios**
 - 5.1. El programa de gestión de datos**
 - 5.2. Conocimientos, competencias y habilidades**
- 6. Nociones teóricas fundamentales**
 - 6.1. El ciclo de vida de los datos**
 - 6.2. Tipología de los datos de investigación**
 - 6.3. El tratamiento de los datos sensibles**
 - 6.4. Los planes de gestión de datos**
 - 6.5. Los repositorios de datos**
 - 6.6. Comunicación y publicación de los datos**
 - 6.7. Los principios FAIR**
 - 6.8. La cita de los datos**
 - 6.9. La formación y actualización profesional**
- 7. Conclusión**
- 8. Bibliografía**

1. PRÓLOGO

Diferentes tendencias están generando transformaciones profundas en los modos de desarrollar investigaciones científicas. La ciencia tradicional se encuentra interpelada por fenómenos como el acceso abierto, el uso intensivo de técnicas computacionales, los datos masivos, la ciencia ciudadana, y los datos abiertos, entre otras. En este escenario, la producción y disponibilidad de datos científicos es creciente y cobra un valor significativo para la propia validación de los resultados. Por este motivo, considerando a los datos como uno de los principales activos de la ciencia, un grupo creciente de agencias de financiamiento e instituciones impulsan normativas, políticas y protocolos para la curaduría, preservación y acceso de los datos. Los datos también son recursos de información que los investigadores necesitan organizar, resguardar y encontrar para poder utilizar. Desde este punto de vista, los datos no difieren en términos generales, de los recursos de información que son el insumo natural de los servicios y productos bibliotecarios. Por otro lado, las bibliotecas académicas y especializadas forman parte del sistema de investigación como proveedoras de servicios relevantes para la comunidad científica y para el público en general. Dada su amplia experiencia en la gestión de información, siendo actores establecidos y reconocidos en sus comunidades, resulta natural que se encuentren llamadas a proveer servicios que promuevan la integridad y disponibilidad de los datos primarios desde que se generan, hasta que se reutilizan por parte de los investigadores.

Describir las maneras en que las bibliotecas se relacionan con la gestión de datos de investigación, e identificar algunas de las competencias y actividades necesarias para prestar servicios de gestión de datos, contribuye a diseñar un marco de referencia que ayude a los bibliotecarios a acompañar la transición que actualmente atraviesa la ciencia. Por otro lado, una mejor comprensión del ciclo de vida de los datos y de otras nociones importantes, contribuirá a fortalecer la función de las bibliotecas y el rol de los bibliotecarios en sus comunidades.

También existe una justificación de índole jurídica normativa, dado que rige en Argentina la Ley 26.899 (promulgada el 13 de noviembre de 2013 y reglamentada el 16 de noviembre de 2016), que crea el Sistema Nacional de Repositorios Institucionales de Acceso Abierto. Esta ley es el marco para el acceso abierto de las publicaciones y datos producidos por investigaciones financiadas con fondos públicos, y establece una serie de obligaciones para las instituciones que componen el Sistema Nacional de Ciencia, Tecnología e Innovación, incluyendo que los bibliotecarios formen parte de los equipos de gestión de los repositorios digitales.

El presente trabajo se propone analizar los principales elementos de la gestión de datos primarios de investigación (GDI). Este análisis permitirá establecer la relación de las bibliotecas académicas y especializadas con la gestión de datos, identificando además, actividades y competencias de los bibliotecarios de datos.

¿Cuáles son los principales elementos de la GDI? ¿Existe una relación entre las bibliotecas académicas y especializadas con la GDI? De existir esa relación, ¿cuáles son las competencias y actividades que desarrollan los bibliotecarios en ese campo? Estas son algunas de las preguntas iniciales que motivan las reflexiones de este trabajo. Partiendo de la premisa de que sí existe aquella relación y que esa relación es, además de natural, positiva para la comunidad académica, se considera que las actividades y las competencias de los perfiles bibliotecarios especializados en datos, contribuyen a consolidar esa relación.

Durante la conducción de este trabajo se realizó una revisión no exhaustiva de la literatura publicada entre 2012-2020 para recuperar bibliografía relevante en Scopus y Google Académico. Los términos de búsqueda utilizados, coincidentes con las variables principales de la investigación, fueron: “*datos abiertos de investigación*”, “*plan de gestión de datos*”, “*gestión de datos de investigación*”, todos los términos se relacionaron en sintaxis de búsquedas que incluyeran “*bibliotecarios*” y/o “*bibliotecas universitarias*”. “Competencias”, fue otra categoría de análisis y se empleó para identificar tanto conocimientos, como capacidades y habilidades, siguiendo la propuesta de Olaz para conceptualizar “competencias” (Olaz, 2011). Se obtuvo un corpus de documentos pertinentes sobre los cuales se realizó una búsqueda manual para identificar nueva bibliografía de interés. Publicaciones oficiales y sitios especializados fueron incluidos en el estudio.

Una propuesta más exhaustiva debería sistematizar las competencias para los bibliotecarios de datos y valorar posibles modificaciones en planes de estudio y/o la generación de ofertas académicas de posgrado en el campo de datos. Sería necesaria una revisión sistemática para barrer todos los estudios publicados en búsqueda de competencias. Dado el incipiente tratamiento y desarrollo del tema en nuestro país, la metodología no incluyó indagaciones a actores institucionales, pero esto será un interesante campo de investigación en el futuro.

2. INTRODUCCIÓN

El presente trabajo propone una aproximación al campo de la gestión de datos de investigación, desde una perspectiva bibliotecológica. Existen elementos conceptuales, actividades y competencias necesarias, para que los bibliotecarios profesionales presten servicios relacionados con datos abiertos en las bibliotecas de instituciones de educación superior e investigación. En este punto, es necesario abordar ciertos conceptos.

En primer lugar, la reflexión teórica acerca del término “competencias” se puede rastrear en pensadores como McClelland que, durante la década de 1970, lo inscribieron dentro contextos productivos para medir el rendimiento laboral de los empleados (Olaz Capitán, Ortíz García, & Sánchez-Mora Molina, 2011; Olaz, 2011). Posteriormente, Bloom (1975) inserta el término en el ámbito educativo al escribir sobre “*enseñanza basada en competencias*”, idea vigente en modelos educativos del presente (Olaz, 2011). Remarcando la relación entre competencias y el ámbito de actuación profesional, para Bung (Tejada & Rodríguez, 2003, p. 10), las competencias son el conjunto de conocimientos, destrezas y

aptitudes que permiten la realización de las actividades profesionales. Olaz entiende *competencia* “como aquel conjunto de conocimientos, capacidades y habilidades, modeladas por el carácter individual o grupal de quien las pone en práctica y, en última instancia, las gestiona.” (Olaz, 2011, p. 590). Según este esquema, los *conocimientos* están formados por elementos teórico-prácticos, formales e informales, que adquiere una persona en relación con su práctica; las *capacidades* se asumen como el potencial innato que el individuo desarrolla según influencias ecológicas y ambientales, y que le permiten gestionar distintas situaciones; las *habilidades* hacen referencia a destrezas que, al ser entrenadas, mejoran la interacción entre la persona y su entorno (Olaz, 2011).

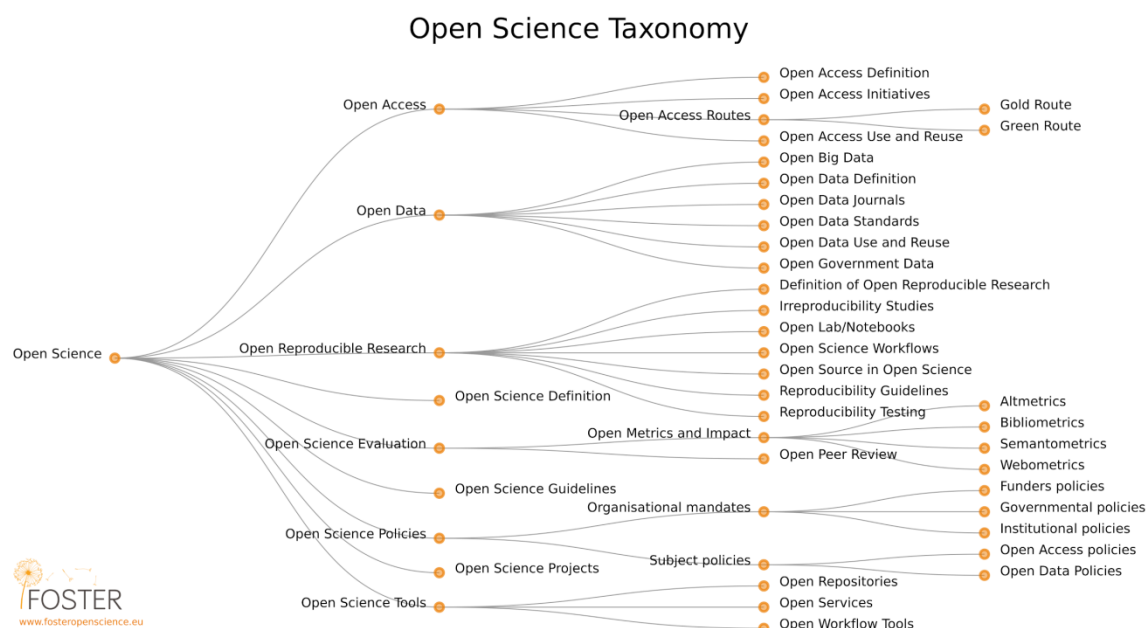
No siendo el propósito central de este trabajo discutir las propuestas precedentes, sino fijar una base conceptual que sirva para anclar las interacciones y prácticas que los bibliotecarios pueden emplear para brindar servicios relacionados con datos, en lo sucesivo se usará el término “competencias”, según el concepto propuesto por Olaz.

Otro concepto que requiere clarificación en el marco de esta investigación es el de datos abiertos de investigación. Sin embargo, para lograr una mayor comprensión, es necesario que previamente se explique qué se entiende por “acceso abierto” y “ciencia abierta”, ambas en relación con los datos abiertos.

El concepto de acceso abierto a la producción científica precede al de *ciencia abierta* (Anglada & Abadal, 2018). La legislación argentina, por medio de la Resolución 753-E/2016, que reglamenta la Ley 26.899 de Repositorios Digitales Institucionales de Acceso Abierto, es clara al definir el modelo de acceso abierto como aquél que permite a cualquier ciudadano “*leer, descargar, copiar, distribuir, imprimir, buscar o enlazar*” de manera gratuita, la producción científico-tecnológica financiada con fondos públicos (considerandos de la Res. 753-E/2016). Idea que está en línea con la definición adoptada por la Budapest Open Access Initiative (2002).

Siguiendo a Anglada y Abadal (2018), en la actualidad la ciencia abierta implica mucho más que el acceso a los documentos que comunican los resultados de una investigación (principalmente artículos científicos, informes y tesis). El modelo de ciencia abierta propuesto por el Proyecto FOSTER de la Unión Europea (ver figura 1), requiere la concurrencia de ocho elementos: open notebooks; datos abiertos; revisión abierta; acceso abierto; software libre; redes sociales académicas; ciencia ciudadana; y recursos educativos abiertos (Anglada & Abadal, 2018). Según el Manual de Capacitación sobre Ciencia Abierta del Proyecto FOSTER, las prácticas de ciencia abierta promueven la colaboración y contribución por medio del acceso gratuito, usando licencias con amplios permisos que permiten la reutilización, redistribución y reproducción de resultados, datos y métodos (Bezjak, S. y otros, 2019).

Figura 1. Taxonomía de la Ciencia Abierta



Se ha mostrado que la propuesta del Proyecto FOSTER incluye los datos abiertos de investigación como elemento de la ciencia abierta. El concepto de “datos abiertos de investigación”, debe ser abordado desde una perspectiva todo-parte. Dicho de otro modo, se debe partir de clarificar el concepto de “dato”, y luego continuar hacia una clase muy particular de datos, que son los denominados “datos de investigación”, para unificarlos finalmente, con la idea de acceso abierto.

En Argentina, el glosario de la Resolución 640-E/2017 (Programa de Datos Abiertos en Ciencia y Tecnología), conceptualiza los datos como: “*el hecho, concepto, instrucción, carácter o conjunto de caracteres, que se expresa por sí mismo, representado de una manera apropiada para que sea comunicado, transmitido o procesado por seres humanos o por medios automáticos, y al cual se le asigna o se les puede asignar un significado.*” Este trabajo se centrará específicamente en los datos de investigación, un tipo de datos generados y/o tratados durante el proceso de investigación científica. Al respecto, el artículo 3° de la Ley N° 26.899 (Repositorios Digitales Institucionales de Acceso Abierto), define dato primario como “*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*”. El glosario de la citada Resolución 753-E/2016, distingue los datos abiertos de los datos primarios, considerando que los datos abiertos son aquellos datos primarios que se han disponibilizado para su acceso abierto. Entonces, de no mediar restricciones de acceso debidamente fundadas, los datos primarios de investigación, también pertenecen a la categoría de “datos públicos” adoptada en la Resolución 640/E-2017.

Los datos generalmente se agrupan formando *Datasets*, *data set* o *conjunto de datos*, lo que hace referencia a una colección de datos “*codificados en una estructura definida, por ejemplo listas, tablas, bases de datos, etc.*” (Glosario, Res. 753-E/2016).

En adelante, este trabajo se referirá a los datos abiertos de investigación en el sentido que marcan las normas citadas precedentemente, y se utilizarán las expresiones “datos primarios abiertos”, “datos científicos abiertos”, “datos abiertos” y “datos abiertos de investigación” como equivalentes.

El acceso abierto a los datos de investigación en el marco de una ciencia crecientemente colaborativa, se hace cada vez más importante y ha sido ampliamente documentado (Tenopir, Birch, & Allard, 2012; Antell K, Foote JB, Turner J, Shults B., 2014). Choudhury (2008) afirma que los datos son una forma de publicación en sí misma. La Declaración de San Francisco sobre la evaluación de la investigación (DORA, Declaration on Research Assessment), considera los conjuntos de datos como un producto importante de la investigación científica, y recomienda que sean parte de la evaluación de la investigación (2013). Todo esto se maximiza, si el valor de los datos se mide en el contexto del Big Data, donde la ciencia se beneficia de métodos de análisis de datos a gran escala.

Dentro del conjunto de conceptos que se abordan en este trabajo, también corresponde explicar qué se entiende por *gestión de datos de investigación*. Pinfield, Cox y Smith la definen como “*una serie de diferentes actividades y procesos asociados con el ciclo de vida de los datos, que involucran el diseño y creación de datos, almacenamiento, seguridad, preservación, recuperación, compartir y reutilizar, todo teniendo en cuenta las capacidades técnicas, consideraciones éticas, cuestiones legales y marcos de gobernanza*” (Pinfield, Cox y Smith, 2014, p. 3). En términos generales, esta idea hace referencia a todas las operaciones tendientes a lograr que los datos primarios implicados en un proceso de investigación, sean almacenados en un sistema que permita preservarlos y recuperarlos para su reutilización en un nuevo ciclo de investigación.

En línea con las condiciones del acceso abierto, los datos abiertos de investigación deberán ser de acceso gratuito para permitir que el público pueda reutilizarlos, remezclarlos y redistribuirlos en los términos que fijen las licencias que tengan asignadas (Bezjak, S. y otros, 2019).

3. La apertura de la ciencia

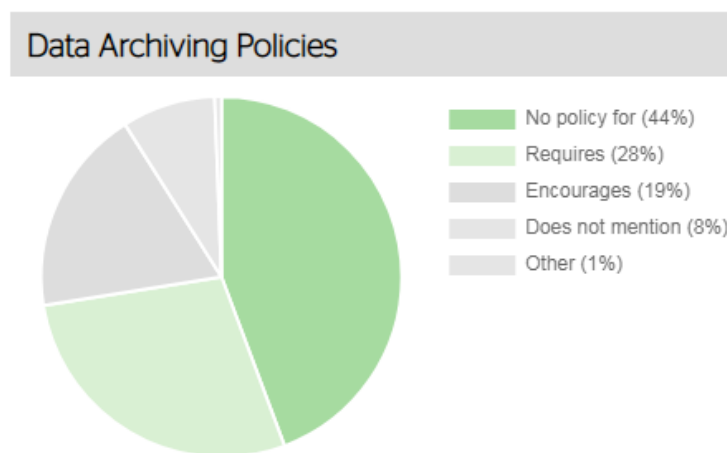
Quizás la iniciativa sobre acceso abierto al conocimiento científico más ambiciosa, sea el *Plan S*, promovido por una coalición de instituciones internacionales con el apoyo de la Unión Europea. Este proyecto se propone lograr que todas las investigaciones financiadas con fondos públicos, se encuentren publicadas en abierto sin ningún tipo de restricción a partir de 2020 (cOAlition S, 2018).

Margaret Mering (2020) define los mandatos de acceso abierto (*OA mandates*), como las políticas adoptadas por las instituciones para que los investigadores aseguren el acceso abierto de sus investigaciones a través de revistas y repositorios de acceso abierto. Estas políticas tienen además, carácter obligatorio (glosario de la Res. 753-E/2016).

Las políticas científicas sobre acceso abierto, consagradas en más de 1050 normativas específicas, no dejan de crecer desde 2005, como se observa en el Registry of Open Access Repository Mandates and Policies (ROARMAP). En este repositorio, se pueden encontrar 54 políticas correspondientes a Sudamérica, de las cuales 7 pertenecen a nuestro país (a la fecha de este trabajo).

Los datos generados por investigaciones financiadas con fondos públicos, están alcanzados por las políticas públicas de acceso abierto. El 63% de las políticas de los organismos financiadores registradas en la base de datos SHERPA JULIET (octubre de 2020), tienen algún tipo de requisito respecto del depósito de datos en acceso abierto (ver figura 2). En Norteamérica, por citar un ejemplo, el National Institutes of Health (NIH), puso en revisión pública, un borrador de política para que los solicitantes de financiamiento presenten sus planes de gestión de datos (National Institutes of Health, 2019). Como se ha mencionado anteriormente, la Argentina tiene una ley nacional conforme la cual toda la producción científica financiada por el Estado, debe ser disponibilizada en acceso abierto a través de repositorios institucionales, lo cual incluye a los datos (Ley N° 26.899).

Figura 2. Políticas sobre archivo de datos en SHERPA JULIET



La importancia que reviste la apertura de la ciencia, reflejada en las políticas mandatorias, se reafirma en distintos consensos multilaterales y compromisos de diversas organizaciones a nivel internacional, de los que aquí sólo haremos una breve referencia:

- ❑ Budapest Open Access Initiative (2002)
- ❑ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003),
- ❑ Bethesda Statement on Open Access Publishing (2003)

- ❑ IFLA Statement on Open Access to Scholarly Literature and Research Documentation (2003)
- ❑ OECD Declaration on Access to Research Data from Public Funding (2004)
- ❑ Salvador Declaration on Open Access: The Developing World Perspective (2005)
- ❑ Ten years on from the Budapest Open Access Initiative: setting the default to open (2012)
- ❑ Science Europe Principles on Open Access to Research Publications (2015),
- ❑ Mallorca Declaration on Open Science (2017)
- ❑ Plan S (2018).
- ❑ The Vienna Declaration on the European Open Science Cloud (2018)
- ❑ The Beijing Declaration on Research Data (2019)
- ❑ Declaration for the Sharing and Opening of Research Data for Sustainable Development (2019)
- ❑ Sorbonne Declaration on Research Data Rights (2020)

4. Beneficios de lo abierto

Bajo este entramado de normas y declaraciones, vale preguntarse cuáles son los beneficios de lo abierto frente a las prácticas tradicionales de publicación a través de costosas licencias de acceso. En un artículo publicado por McKiernan y otros (2016), se presenta evidencia cuantitativa indicando que las prácticas de acceso abierto revisten cinco importantes beneficios para el desarrollo de la carrera de investigación:

- ❑ Aumentar el número de citas obtenidas;
- ❑ Ganar visibilidad a través de medios de comunicación y redes sociales;
- ❑ Llegar a potenciales colaboradores (dado que el acceso abierto facilita el descubrimiento de las investigaciones y la colaboración);
- ❑ Mejorar las oportunidades laborales;
- ❑ Acceder a más opciones de financiamiento

Saliendo de la esfera puramente académica, los beneficios de la ciencia abierta también impactan en la vida social. La estrecha relación entre acceso abierto, ciencia, bienestar, innovación, desarrollo social y económico, también es un argumento valioso en favor del acceso abierto. La Resolución 640-E/2017, considera que el acceso público y gratuito a la producción científica, contribuye a crear una cultura de colaboración y a desarrollar el patrimonio cultural y social del país. El derecho de los ciudadanos a conocer en qué se invierten los fondos públicos destinados a la ciencia y acceder libremente a ellos, comporta un beneficio notable (McKiernan y otros, 2016), que ayuda a la transparencia y la participación ciudadana.

Estas ventajas también se aplican a la apertura de los datos de investigación. Diferentes estudios afirman que depositar los datos en acceso abierto promueve las citas y la visibilidad de la investigación, al tiempo que permite validarlos (McKiernan y otros, 2016; Melero, R.;

Hernández-San-Miguel, J., 2014; Pinfield, Cox y Smith, 2014). Abrir los datos fortalece el proceso de verificación de los resultados al permitir el escrutinio de otros investigadores sobre los datos primarios (Bezjak, S. y otros, 2019).

En base a lo expuesto, se puede esperar que el acceso a los datos primarios por parte de toda la comunidad científica, brinde mayor transparencia al proceso de creación de conocimiento. La disposición en abierto de los datos de investigación comporta cumplir con los mandatos institucionales, fortalecer la credibilidad y facilitar su reutilización, lo que redundará en el uso eficiente de los recursos técnicos y económicos (Melero, R.; Hernández-San-Miguel, J., 2014).

5. Los datos, las bibliotecas y los bibliotecarios

La relación entre los datos de investigación y los bibliotecarios de bibliotecas universitarias y centros de investigación, es un campo de reflexión que se encuentra en una fase temprana (Faniel & Connaway, 2018; Brochu & Burns, 2019). Con las primeras discusiones hacia 2008-2009, para 2014 las bibliotecas buscaban ser reconocidas como un actor institucional válido en la gestión de datos, necesidad que Pinfield, Cox y Smith (2014), identifican como la formación de un “*factor jurisdiccional*” de las bibliotecas.

Gran parte de los estudios que tratan el tema consisten en encuestas realizadas a bibliotecarios de bibliotecas universitarias (Brown, 2010; Antell K, Foote JB, Turner J, Shults B., 2014; Tenopir et al., 2015; Matusiak, K. y Sposito, F., 2017; Tenopir et al., 2017), estudios de caso (Choudhury GS., 2008; Mullins, JL., 2012; Witt M., 2012; Hswe y Holt, 2011), o entrevistas. De cualquier modo, provienen del ámbito anglosajón. Los trabajos de Faniel (2018) y de Pinfield, Cox y Smith (2014), se basan en entrevistas a bibliotecarios profesionales de Reino Unido, Estados Unidos, Canadá, Europa, Australia y Nueva Zelanda. Argentina cuenta con reglamentaciones vigentes sobre el acceso a los resultados de la investigación científica, que incluyen a los bibliotecarios en los equipos técnicos encargados de la gestión de repositorios digitales (Ley 26.899; Resolución 753-E/2016). Las normativas constituyen una responsabilidad tanto para los investigadores, como para las instituciones del Sistema Nacional de Ciencia y Tecnología donde los bibliotecarios prestan servicio.

Las reglamentaciones vigentes en los distintos países conforman una fuerza poderosa que influye en el campo de las bibliotecas (Antell K, Foote JB, Turner J, Shults B., 2014; Matusiak, K. y Sposito, F., 2017). En una perspectiva más amplia e integradora y, sin perder de vista los matices de su especificidad, se podría considerar que el acceso a los datos primarios forma parte del campo del acceso a la información en general y, en la medida en que el Estado participe de su financiamiento y creación, del acceso a la información gubernamental. Siendo así, el acceso a los datos abiertos en Argentina debe considerarse como parte de un esquema normativo amplio en favor del acceso a la información pública: las ya citadas ley 26.899 y su resolución reglamentaria, N° 753-E/2016; Ley 27.275 que regula el ejercicio del derecho de acceso a la información pública, y su decreto reglamentario No. 266/17; Decreto 117/2016 por el cual Ministerios, Secretarías y organismos

desconcentrados y descentralizados dependientes del Poder Ejecutivo Nacional deben tener un Plan de Apertura de Datos; Resolución 11/2016 por la cual Argentina adhiere a la Carta Internacional de Datos Abiertos; Programa de Datos Abiertos en Ciencia y Tecnología (Resolución 640-E/2017, para coordinar, entre otros objetivos, las distintas iniciativas estatales sobre datos abiertos); Resolución 538/2013 que reglamenta el Sistema Nacional de Datos Públicos (SINDAP); y Decreto 1172/2003 de Régimen de Acceso a la Información Pública.

Además, la incorporación de los datos de investigación en repositorios de acceso abierto se está consolidando como mandato para la comunidad científica global. A nivel internacional, distintos organismos exigen a los grupos de investigación financiados con fondos públicos, que incluyan un plan formal para la gestión de datos en las solicitudes de financiamiento (Antell K, Foote JB, Turner J, Shults B., 2014; Dietrich D, Adamus T, Miner A, Steinhart G, 2012; Matusiak, K. y Sposito, F., 2017). Hacia 2018, un informe desarrollado de manera conjunta entre el Digital Curation Centre y SPARC Europe, mostraba que 11 de los 28 miembros de la Unión Europea tenían políticas nacionales relacionadas con los datos de investigación (Digital Curation Centre, 2018). La última versión del informe mantiene la cantidad de políticas nacionales, pero otros 15 estados están promoviendo diferentes estrategias relacionadas (Digital Curation Centre, 2020). Esta situación no resulta desconocida para los bibliotecarios en las instituciones de investigación de los países más desarrollados (Antell K, Foote JB, Turner J, Shults B., 2014.). En ese contexto, la exigencia también se consagra a nivel nacional en el art. 3° de la ley 26.899 y en la Resolución 753-E/2016.

Brochu y Burns (2019) concluyen que las políticas que mencionan a los bibliotecarios en los planes de gestión de datos están creciendo en todo el mundo. En nuestro país, se ha señalado que la legislación sobre repositorios institucionales contempla los datos primarios de investigación y mencionan a las bibliotecas y a los bibliotecarios (art. 9 Res. 753-E/2016).

Pinfield, Cox y Smith (2014), afirman que las bibliotecas desarrollan un rol importante en este campo de acción, e identifican tres áreas de trabajo asumidas de manera temprana: desarrollo de políticas y guías, desarrollo de infraestructura tecnológica y servicios de apoyo. Nitecki y Davis (2019), consideran que los servicios bibliotecarios viven un cambio importante hacia servicios de gestión y difusión de datos, lo cual expande el rol de los bibliotecarios. Brochu y Burns (2019) afirman que los bibliotecarios tienen una función muy valiosa y su implicación temprana dentro del proceso promueve una mejor gestión y cuidado de los datos.

5.1 El programa de gestión de datos: oportunidades y desafíos para las bibliotecas

Existe un nivel macro que las bibliotecas deben considerar para comprender el alcance de las oportunidades y la profundidad de los desafíos intrínsecos de la gestión de datos. Todo el proceso se instala en un contexto institucional diverso y complejo, dentro del cual la biblioteca está inserta. Los distintos actores tienen perspectivas, modos de hacer, e intereses

propios que deben ser armonizados a través de un esfuerzo colaborativo, conducente a lograr una adecuada gestión de los datos de investigación. En este sentido, la creación de un programa institucional de gestión de datos, será el marco que organice todos los intereses, actividades y recursos para el logro de los objetivos. Pinfield, Cox y Smith (2014), identifican seis componentes de un programa de gestión de datos: estrategias, políticas, guías, procesos, tecnología y servicios:

- las **estrategias** hacen referencia a la visión general de la institución para los datos que produce;
- las **políticas y guías** permiten operacionalizar las estrategias;
- los **procesos** guían todas las actividades dentro del ciclo de vida de los datos;
- las **tecnologías** son necesarias para desarrollar las actividades (por ej. los repositorios de datos);
- los **servicios**, destinados tanto a los investigadores, como al usuario final.

Continuando con su explicación, describen distintos factores que influyen sobre la relación entre las bibliotecas y la gestión de datos: almacenamiento (capacidad de la infraestructura para almacenar grandes cantidades de datos); seguridad (especialmente de aquellos datos sensibles y confidenciales); preservación (capacidad y actividades para asegurar el resguardo a largo plazo); capacidad para cumplir con políticas y mandatos; calidad (relacionado con la capacidad de los datos abiertos para promover la verificación y reproducibilidad de los estudios); acceso y reutilización (apertura de los datos); factor jurisdiccional (reconocimiento de la biblioteca como un actor institucional en la materia).

¿Qué lugar ocupan las bibliotecas en estos programas? ¿Cuáles son las áreas de mayor incumbencia bibliotecaria? Hacia 2014 las bibliotecas estaban implicadas en al menos tres áreas de desarrollo de un programa de gestión de datos: cumplimiento de requisitos y políticas; preservación, y acceso a datos (Pinfield, Cox y Smith, 2014).

En cuanto al desarrollo de políticas, guías y procedimientos, las bibliotecas y los bibliotecarios tienen una larga tradición en este camino. Políticas para el desarrollo de las colecciones, políticas para la descripción documental, políticas para el intercambio de recursos, etc. forman parte de una cotidianeidad bibliotecaria a la cual hay que sumar la estandarización de procesos y la utilización de normas internacionales, todo lo cual también se señala en la literatura relevada (Antell y otros, 2014; Nitecki y Davis, 2019). Los testimonios recogidos en los estudios citados verifican la implicación de los bibliotecarios en grupos de trabajo sobre políticas institucionales de datos (Pinfield, Cox y Smith, 2014; Matusiak, K. y Sposito, F., 2017)

La experiencia verificada (Hswe y Holt, 2011; Antell y otros, 2014; Brochu y Burns, 2019; Pinfield, Cox y Smith, 2014), evidencia grandes avances de las bibliotecas en orden a garantizar la preservación y el acceso de los datos no sólo en cuanto a la documentación antes mencionada (políticas, procedimientos, guías), sino también en infraestructura tecnológica

(repositorios institucionales) y servicios especializados (metadatos, alfabetización de datos, referencia y orientación sobre datos y gestión datos de investigación).

Ante este escenario, corresponde a la comunidad bibliotecaria asumir el desafío y multiplicar esas experiencias, brindando apoyo a los investigadores para que gestionen los datos de investigación en un marco estandarizado y seguro. El estudio de Antell y otros (2014), destaca que trabajar en contacto más cercano con los investigadores, ofrecer acceso a datos y desarrollar nuevas habilidades, supone una oportunidad para los bibliotecarios. Algunas bibliotecas universitarias marcan antecedentes favorables en este camino, como las bibliotecas de Purdue University, que impulsaron la creación de un grupo de trabajo (PURR Working Group) para que los bibliotecarios actúen en colaboración con investigadores y otros actores institucionales en la curaduría de datos (Witt, 2012). Para reforzar esta idea, son interesantes los emergentes del estudio de Pinfield, Cox y Smith (2014), con testimonios de bibliotecarios implicados en actividades de gestión de datos, que destacan la necesidad de cooperación con departamentos como sistemas, asesoría legal, ciencia y tecnología, y con los propios investigadores. El mismo estudio muestra que existen experiencias en que las bibliotecas asumen además, la tarea de coordinar a los distintos actores institucionales. Parece existir consenso sobre la interdisciplinariedad y la colaboración que requiere este área de trabajo (Hswe y Holt, 2011; Brown, 2010; Antell y otros, 2014; Brochu y Burns, 2019). En este sentido, el trabajo de Tripaldi (2018) enfatiza que la legislación argentina incluye de manera explícita a los bibliotecarios en los equipos de trabajo de los repositorios digitales, y discute los alcances de la interdisciplinariedad implícita en el texto legislativo. Establecer la naturaleza de estas relaciones dentro de este campo de acción merece mayores indagaciones.

La gestión de datos de investigación implica oportunidades y desafíos para los bibliotecarios. Las oportunidades están dadas por la posibilidad de retomar el contacto directo con los investigadores y todo lo que ello puede implicar en términos de compartir objetivos, enriquecer la visión que ambas partes tienen acerca de la otra, fortalecer el rol de las bibliotecas como unidades de valor para la comunidad, etc. Pero también hay oportunidades en la experiencia, formación, y saber hacer de los bibliotecarios, que constituyen competencias inherentes a su formación profesional. Los bibliotecarios ya poseen habilidades tradicionales para organizar información, usar estándares, asignar metadatos, enseñar a buscar, planificar, etc. (Antell y otros, 2014; Nitecki y Davis, 2019); además las bibliotecas poseen la longevidad e infraestructura necesarias para asumir la tarea (Heidorn, 2011). El desconocimiento de los investigadores sobre dónde depositar los datos, también abre un campo de oportunidades, sobre todo para los bibliotecarios con las competencias adecuadas para asistir y guiar a los investigadores en la gestión de sus datos (Dietrich, 2012).

5.2 Conocimientos, competencias, habilidades

Es reconocido el rol de las bibliotecas universitarias y especializadas como proveedoras de servicios para el aprendizaje y la investigación. Estos servicios se basan en ciertos recursos e infraestructuras, pero el activo más importante de las bibliotecas son los profesionales y las competencias que les permiten relacionar todos los activos para brindar servicios. En el caso

de la gestión de datos, este principio se repite. Por lo tanto, para responder a una de las preguntas conductoras de este trabajo, es necesario identificar los conocimientos, capacidades y habilidades que conforman las competencias necesarias para implicarse en la gestión de datos de investigación desde las bibliotecas.

Pinfield, Cox y Smith (2014), señalan que las habilidades son uno de los factores que influyen sobre el desarrollo de este campo de acción. En el trabajo de Antell y otros (2014), la mayor parte de los bibliotecarios consultados consideran que necesitarán habilidades específicas, entre las que mencionan: conocimiento del ciclo de vida de los datos, habilidades informáticas, experiencia en el proceso de solicitud de financiamiento y conocimiento del campo disciplinar. Brown (2010) asigna gran potencial para los bibliotecarios que se sumen a proyectos colaborativos de curaduría de datos junto a investigadores. Aunque, como señalan Brochu y Burns (2019), la participación de los bibliotecarios en el proceso dependerá de las decisiones del gobierno de las instituciones. Esta última afirmación se puede aplicar en ciertos contextos institucionales donde las políticas no explicitan el rol de la biblioteca como parte del proceso de gestión de los datos, lo cual no debería suceder en países como Argentina, donde la normativa nacional es clara al respecto.

La lectura ha permitido identificar un gran número de términos en inglés que denominan las tareas relacionadas con la gestión de datos. Una correcta delimitación del alcance de cada uno merece una propuesta de investigación independiente: “data management support services”, “data management assistance”, “data curation”, “data literacy” (Antell K, Foote JB, Turner J, Shults B., 2014)

Uno de los términos más repetidos en la literatura y que aparece relacionado con los bibliotecarios, es curaduría de datos (data curation). El glosario de la Res. 753-E/2016 define curaduría de datos como un conjunto de actividades para mantener, preservar y agregar valor a los datos de investigación durante todo el ciclo de vida. De este modo, la curaduría es un proceso completo y continuo, no una actividad puntual.

Una de las actividades fundacionales que los bibliotecarios tendrán que emprender en las organizaciones, será conocer muy bien a su comunidad de investigadores, específicamente en lo concerniente a las particularidades de los datos generados en los procesos de investigación y a los modos en que ellos almacenan, comparten y/o preservan esos datos. Esto les permitirá tener un mejor entendimiento de las necesidades que surjan en el marco de los planes de gestión de datos. Vale destacar dos iniciativas interesantes referidas en la bibliografía, emprendidas por las bibliotecas de Georgia Tech y de la University of Houston, orientadas a conocer las prácticas de gestión de datos de sus investigadores (Antell K. y otros, 2014). Todo conocimiento de las necesidades de los investigadores se verá reforzado por el contacto directo. Los bibliotecarios deben participar y proyectarse en el ciclo de vida de la investigación de manera temprana, establecer contacto con los grupos interesados e implicarse en el desarrollo de estándares de metadatos (Luce, 2008). Habilidades blandas, como las comunicativas y de relación, son importantes para generar colaboración con los principales actores institucionales (Pinfield, Cox y Smith, 2014).

Faniel (2018), postula cinco factores que influyen la capacidad de los bibliotecarios para atender las necesidades de los investigadores en torno a la gestión de sus datos de investigación. Estos factores pueden facilitar o dificultar ese objetivo: 1) recursos técnicos; 2) recursos humanos; 3) percepción de los investigadores acerca de la biblioteca; 4) apoyo de los líderes institucionales y 5) comunicación, coordinación y colaboración.

En las consultas realizadas, Antell y otros (2014) notan la existencia de una sensación dual entre los bibliotecarios. Por un lado, de optimismo porque las habilidades tradicionales se aplican al nuevo campo y, por otro lado, de incertidumbre en torno al rol de las bibliotecas y las nuevas habilidades. Brochu y Burns (2019) apuntan que, entre algunos bibliotecarios de universidades, existe la idea de que asistir a los estudiantes es más importante que asistir a los investigadores.

Nitecki y Davis (2019), consideran que los servicios bibliotecarios viven un cambio importante hacia la gestión y difusión de datos. Proponen cuatro supuestos relacionados entre sí para comprender las oportunidades que surgen: 1) el proceso de investigación incluye el intercambio y reutilización de datos; 2), la gestión de datos abiertos de investigación expande el rol de los bibliotecarios; 3) el nuevo rol de los bibliotecarios contribuye a empoderar a otros; 4) las alianzas fuera de la biblioteca permiten que los bibliotecarios construyan un rol de liderazgo. En este sentido, agregan que los bibliotecarios deben ser emprendedores y proactivos en la transición hacia los servicios de datos, desarrollando un nuevo rol que denominan “*intelectual entrepreneur*”, que implica avanzar desde el rol de servicio de apoyo, hacia una asociación con los investigadores para promover el cambio organizacional y el progreso de la ciencia.

No hay una propuesta unívoca sobre cuáles son las competencias necesarias para cubrir las áreas de carencia (Antell y otros, 2014), los documentos son ricos en denominaciones sobre actividades, saberes, actitudes y servicios.

Retomando a Olaz (2011), en el núcleo de una competencia existe un conjunto de conocimientos, capacidades y habilidades; siguiendo su esquema, los *conocimientos* están formados por elementos teórico-prácticos, formales e informales, que adquiere una persona en relación con su práctica. A la luz de esta propuesta, a continuación se reúnen las capacidades, conocimientos y habilidades identificadas en los documentos y que están relacionadas con la gestión de datos desde el punto de vista bibliotecario. Se intentan organizar algunas de las unidades que, a juicio de este trabajo, pueden dar forma a la identidad del bibliotecario de datos. Se las agrupa en dos categorías: “tradicionales”, aquellos elementos teórico-prácticos que los bibliotecarios ya asumen como parte de su identidad profesional; y “nuevas”, elementos que deben adquirir para complementar el perfil:

Tradicional
Organización de la información (Antell y otros, 2014; Nitecki y Davis, 2019)
Aplicación de estándares y metadatos (Antell y otros, 2014).
Conocimientos sobre acceso abierto y estandarización.
Conciencia sobre la preservación de la información.
Provisión de información (Antell y otros, 2014)
Conocimiento sobre el proceso de investigación (Antell y otros, 2014)
Conocimiento sobre el proceso de la comunicación científica.
Habilidad para estar en contacto con docentes e investigadores (Antell y otros, 2014)
Selección de recursos apropiados (Antell y otros, 2014)
Formación de usuarios (Antell y otros, 2014)
Conocimiento sobre bases de datos (Schmidt, B. y Shearer, K., 2016)
Formación y actualización profesional continua (Antell y otros, 2014)

Nuevas:
Gestión y curaduría de datos. Incluye: selección de contenidos para el repositorio, descubrir, reutilizar y recopilar (Antell y otros, 2014; Brochu y Burns, 2019; Pinfield, Cox y Smith, 2014; Matusiak, K. y Sposito, F., 2017); conocimientos sobre estándares y esquemas de metadatos, formatos de datos, ontologías, identificadores persistentes, plataformas de repositorios, citación de datos, licencias de uso y propiedad intelectual (Schmidt, B. y Shearer, K., 2016; Matusiak, K. y Sposito, F., 2017)
Conocimiento de las nuevas tendencias de la comunicación científica. Incluye comunicación y publicación de datos (Garitano y Carlson, 2009; Schmidt, B. y Shearer, K., 2016)
Alfabetización de datos, referencia, orientación, formación y apoyo sobre gestión y descubrimiento datos de investigación (Hswe y Holt, 2011;Antell y otros, 2014; Brochu y

Burns, 2019; Pinfield, Cox y Smith, 2014; Schmidt, B. y Shearer, K., 2016; Matusiak, K. y Sposito, F., 2017)
Promoción, advocacy, enlace, comunicación. Incluye conocimiento sobre buenas prácticas de gestión, mandatos, regulaciones y requisitos de organismos financiadores (Antell y otros, 2014; Pinfield, Cox y Smith, 2014; Schmidt, B. y Shearer, K., 2016; Matusiak, K. y Sposito, F., 2017)
Conocimientos y habilidades en preservación digital (Schmidt, B. y Shearer, K., 2016)
Administración de repositorios (Antell y otros, 2014)
Conocimientos sobre los principios FAIR.
Conocimiento sobre planes de gestión de datos y herramientas para su desarrollo (Antell y otros, 2014; Schmidt, B. y Shearer, K., 2016; Matusiak, K. y Sposito, F., 2017)
Conocimiento y capacidad para desarrollar políticas, protocolos e infraestructuras para la descripción, descubrimiento, recuperación y cita de datos (Nitecki y Davis, 2019; Pinfield, Cox y Smith, 2014; Schmidt, B. y Shearer, K., 2016; Matusiak, K. y Sposito, F., 2017)
Conocimiento del ciclo de vida de los datos (Antell y otros, 2014; Nitecki y Davis, 2019), y tipologías de datos.
Conocimientos sobre el área temática específica y sus prácticas de investigación (Antell y otros, 2014; Schmidt, B. y Shearer, K., 2016)
Trabajo en equipo y comunicación (Antell y otros, 2014)
Conocimientos informáticos (Antell y otros, 2014)
Conocimientos sobre técnicas y herramientas para el análisis de datos (Schmidt, B. y Shearer, K., 2016)
Conocimiento sobre aspectos legales (Antell y otros, 2014); cuestiones éticas y confidencialidad de los datos.

La tabla precedente no propone un universo completo y exhaustivo, sino más bien una fotografía de ciertas regularidades observadas en la literatura seleccionada. La categorización que elaboran Schmidt, B. y Shearer, K. (2016), facilita una clasificación de los servicios de las bibliotecas en tres grupos generales:

- A) Servicios para proporcionar acceso a datos.
- B) Servicios para ayudar a los investigadores y estudiantes en la gestión de sus datos.
- C) Servicios para gestionar colecciones de datos.

Matusiak, K. y Sposito, F. (2017), también identifican casos en que la biblioteca se encarga de la revisión de calidad de los datasets depositados: metadatos, formatos de archivos y documentación.

6. Nociones teóricas fundamentales

Estos servicios implican el desarrollo de diversas tareas, y el conocimiento de diferentes contenidos teóricos. Algunos de esos contenidos, que se presentan en el siguiente apartado como nociones fundamentales, fueron mencionados someramente y requieren un desarrollo más detallado al ser emergentes claros del análisis bibliográfico.

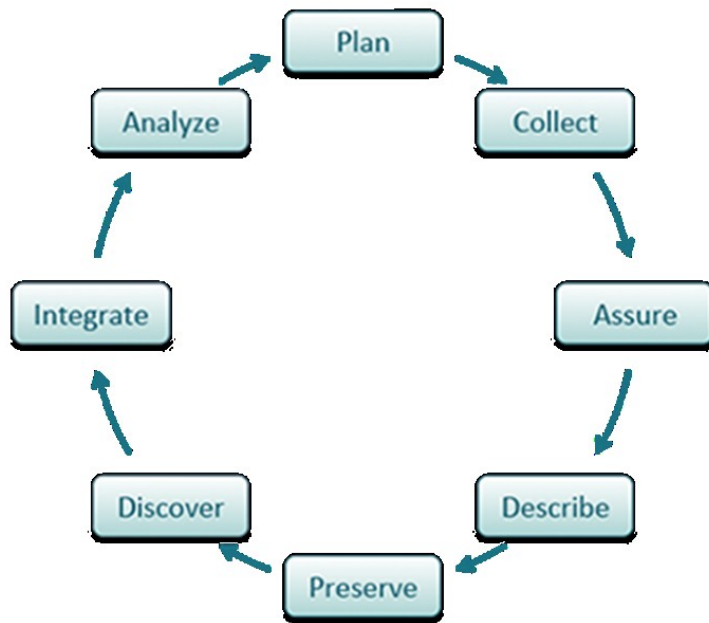
6.1 El ciclo de vida de los datos:

Un ciclo de vida es un modelo que permite identificar de manera ordenada las distintas fases de tratamiento de los datos de investigación desde su creación, hasta su reutilización. Existen diferentes propuestas del ciclo de vida, pero su importancia para el diseño de servicios de datos lo convierte en un modelo vital, ya que permite visualizar las actividades que desarrollan los usuarios de datos y pensar maneras para asistirlos (Guss, S., 2016).

Brochu y Burns (2019), explican un diseño desarrollado por la University of Queensland que consta de seis fases: 1) creación, 2) procesamiento, 3) análisis, 4) preservación, 5) descubrimiento y 6) reutilización. El ciclo se inicia cuando los investigadores recopilan y generan datos. Luego, en la fase de procesamiento, los investigadores trabajan sobre los datos recopilados y generados. Los procesos aplicados sobre los datos y las conclusiones obtenidas se enmarcan como análisis de los datos. La preservación contempla el formato de guarda de los datos, su localización, y el tiempo que estarán disponibles. La fase de descubrimiento se centra en la disponibilidad de los datos y responde las preguntas sobre quién podrá acceder, dónde lo hará o a través de qué herramientas, y cuáles serán las características de ese acceso. En la fase de reutilización se debe garantizar el acceso futuro de los datos para otros investigadores.

El ciclo de vida de los datos propuesto por la Data Observation Network for Earth (DataOne), una colaboración de organizaciones que promueve el acceso a datos de investigación sobre la vida y el medio ambiente, está compuesto por ocho etapas y es aplicable a cualquier área científica (ver figura 3): planificar (describir los datos que se recopilarán, cómo se gestionarán y cómo se harán accesibles), recolectar (los datos son guardados o generados en formato digital), asegurar (controlar y revisar la calidad de los datos), describir (los datos se describen a través de metadatos, utilizando los estándares adecuados), preservar (los datos se depositan en un repositorio o sistema apropiado para su conservación a largo plazo), descubrir (los datos y sus metadatos se encuentran y recuperan), integrar (datos de distintos proyectos se pueden combinar para formar un nuevo *set* (agrupación o conjunto de datos), listo para ser analizado por los investigadores), y analizar (los datos recuperados se analizan en el marco de un nuevo proyecto de investigación).

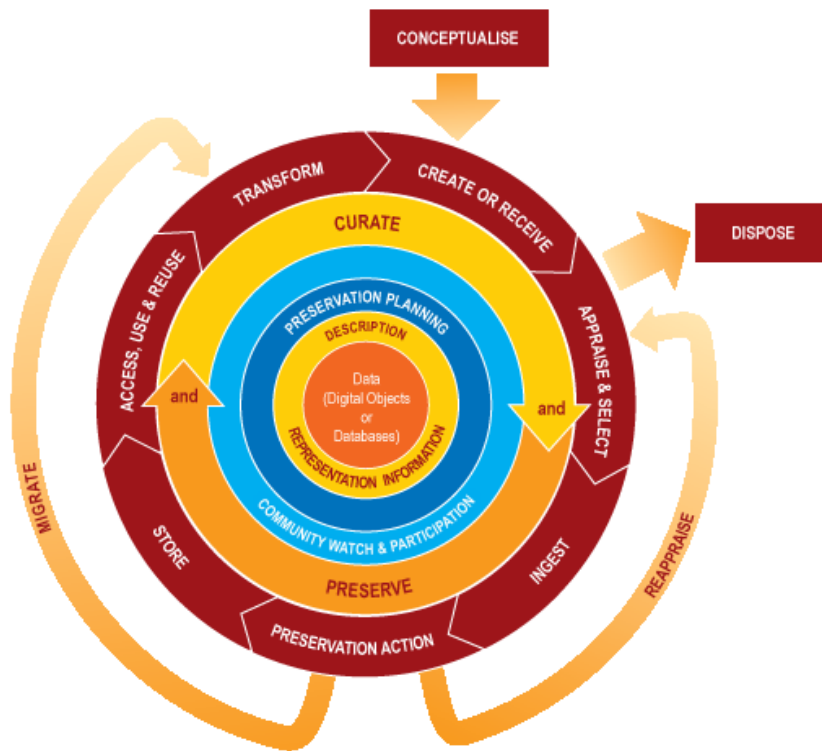
Figura 3. Modelo del ciclo de vida de DataOne



El trabajo de Higgins (2007) identifica otros modelos provenientes de distintas áreas disciplinares. Un modelo extendido es el Curation Lifecycle Model (ver figura 4), propuesto por el Digital Curation Centre (DCC) (Higgins, 2008), como modelo para comprender el ciclo de vida de los datos. Esta propuesta ilustra los flujos de trabajo y roles relacionados con el proceso de gestión de datos. La finalidad de esta herramienta es guiar a los actores para lograr la correcta gestión, curaduría y preservación de datos. Higgins explica la fortaleza de un modelo genérico, más indicativo que exhaustivo, lo que le permite ser adaptado a las necesidades de los distintos campos que trabajan con contenidos digitales y complementar otros estándares que fueran necesarios. Siguiendo la explicación de la autora, y dado que los datos de investigación deben ser identificados como un activo digital, la propuesta del ciclo de vida puede ser aplicada para:

- ❑ definir responsabilidades y roles dentro del proceso;
- ❑ desarrollar estándares y tecnologías;
- ❑ elaborar la documentación de procedimientos y políticas;
- ❑ identificar acciones de curaduría;
- ❑ capacitar, como material didáctico para la formación de los perfiles implicados en el proceso de curaduría y preservación;
- ❑ desarrollar metodologías de trabajo a nivel institucional.

Figura 4. DCC Lifecycle Model



Con los datos en el centro, una serie de anillos se van desplegando hacia el exterior, indicando las distintas acciones de un proceso cuya finalidad, como se explicó anteriormente, es asegurar la integridad y disponibilidad de los datos.

A continuación se traducen las acciones comprendidas dentro del modelo del DCC en base al trabajo citado de Higgins (2008):

Acciones durante todo el ciclo de vida:

- ***Descripción e Información de Representación:*** aplicar estándares de metadatos que aseguren la adecuada descripción del recurso, representación que permita comprender y representar el recurso digital.
- ***Planificación de la Preservación:*** planificar la preservación de los activos digitales durante todo el ciclo de vida.
- ***Vigilancia Comunitaria y Participación:*** Monitorear las actividades y participar en el desarrollo de estándares compartidos, herramientas y software.

- **Curar y Preservar:** Emprender acciones para promover la curaduría y preservación a lo largo de todo el ciclo.

Acciones secuenciales:

- **Conceptualizar:** Explicar y planificar cómo se crearán los datos, incluyendo el método de recolección y opciones de almacenamiento.
- **Crear y Recibir:** Esta acción comprende dos operaciones.
- **Crear datos,** de cualquier clase (administrativos, descriptivos, etc.), incluyendo los metadatos y los metadatos de preservación.
- **Recibir datos** de otros creadores o fuentes como repositorios o centros de datos, respetando las políticas sobre recolección y asignando los metadatos que fueran necesarios.
- **Evaluar y Seleccionar:** Conforme a las políticas y aspectos legales que apliquen, evaluar y seleccionar aquellos datos que serán curados y preservados para el largo plazo.
- **Ingesta:** Conforme a las políticas y aspectos legales que apliquen, depositar los datos en un repositorio, centro de datos u otro sistema.
- **Preservar:** Tomar acciones que aseguren la preservación a largo plazo garantizando que los datos preserven su integridad para que sean auténticos, confiables y reutilizables. Estas acciones incluirán: limpieza de datos, validación, asignación de metadatos de preservación no incluidos en otra etapa, asignación de información de representación y estructuras de datos o formatos de archivo aceptables.
- **Almacenar:** Almacenar los datos de manera segura y estandarizada.
- **Acceso, Uso y Reutilización:** Asegurar que los datos sean accesibles para todos los usuarios en todo momento.
- **Transformación:** Crear nuevos datos a partir de los originales (por migración de formato, creación de un nuevo subconjunto (subset), por selección o consulta, derivando resultados nuevos para su publicación, etc.)

Acciones ocasionales:

- **Desechar o destruir:** En función de políticas documentadas y/o requisitos legales, ciertos datos no son elegibles para su preservación a largo plazo y pueden ser transferidos a otro repositorio o destruidos.
- **Reevaluar:** Si se detectan faltas en el proceso de validación, los datos se devuelven para nueva evaluación y re-selección.

- **Migrar:** Migrar los datos a un formato diferente es un componente importante de la preservación. La migración se puede realizar por un cambio del sistema de almacenamiento o para asegurar la disponibilidad de los datos ante la obsolescencia de hardware o el software.

Pinfield, Cox, Smith (2014) observan que el modelo del DCC está más centrado en la curaduría de datos que en actividades de apoyo, y elabora un diseño diferente, basado en un concepto sencillo que integra los elementos identificados para un programa de gestión de datos a nivel institucional y que fueron expuestos anteriormente.

6.2 Tipología de los datos de investigación

La diversidad de los datos de investigación es tan amplia como la diversidad de campos de investigación, objetos y metodologías de estudio. Por este motivo, algunos autores plantean la necesidad de investigar las características de los datos y de los procesos de investigación que se desarrollan en una comunidad académica específica. La elaboración de perfiles (“data curation profiles”, Witt et al., 2009), o entrevistas con los investigadores (“data interview”, Carlson, J., 2012; Read, K. et al., 2019), parecen técnicas que se inscriben en la tradición bibliotecaria de los estudios de usuarios y las entrevistas de referencia. Estos abordajes también permiten conocer las necesidades de curaduría de datos desde la perspectiva de quienes los crean y utilizan.

La Res. 753-E/2016 indica que los datos de investigación se pueden clasificar en tres categorías: observacionales, experimentales y computacionales. Esta reglamentación también identifica elementos que forman parte de la investigación, pero no se consideran datos primarios: anotaciones de laboratorio, análisis preliminares, objetos físicos como “*muestras, cepas de bacterias, animales de ensayo, vasijas, especímenes, etc.*” (Glosario de la Res.), ni “*borradores de artículos científicos, [...] comunicaciones con colegas, etc.*” (Art. 16).

Los datos de investigación cubren una extensa diversidad que puede incluir: cálculos a gran escala derivados del procesamiento computacional, códigos de programación, datos biológicos, archivos de secuencias de ADN, mediciones, resultados de distintos tipos de experimentos, registros sonoros de entrevistas, videgrabaciones, tablas, archivos CSV, simulaciones, imágenes, etc.

DMPTool, un servicio diseñado por la California Digital Library destinado a los investigadores, adopta cuatro categorías para organizar los datos de investigación según la fuente, el formato, la estabilidad y el volumen (University of California Curation Center DMPTool v3.0.0, 2020):

Según la fuente de los datos	Según el formato de los datos	Según la estabilidad de los datos	Según el volumen de los datos
*Observacional *Experimental *Simulación *Derivada / compilada	*Texto *Numérico *Audiovisual *Modelos, códigos *Formatos específicos (p. ej. FITS en astronomía, CIF en química); formatos generados por instrumentos.	*Conjuntos fijos (no cambian) * Conjuntos en crecimiento (nuevos datos se suman al conjunto sin cambiar los anteriores) * Conjuntos revisables (nuevos datos se suman y los anteriores pueden cambiar o ser borrados)	¿Cuántos datos producirá la investigación? Esto dependerá, por ejemplo, de la fuente y formato de los datos.

6.3 El tratamiento de datos sensibles. Consideraciones legales y éticas.

Muchas investigaciones requieren utilizar datos de las personas (condiciones clínicas, ubicación, historia familiar y social, imágenes, datos personales, etc.) otras, por ejemplo, presentan resultados que pueden ser explotados económicamente, o tienen relevancia para la seguridad de un país. Para proteger estos datos del acceso público, las políticas y normativas reconocen estos derechos e intereses a través de mecanismos de protección o exclusión por el cual no son puestos en acceso público, o se le aplican restricciones. En este sentido, los planes de gestión de datos deben contemplar el tratamiento que se realizará con estos datos.

En Argentina, la Ley de Ciencia, Tecnología e Innovación, establece que se debe respetar la privacidad e intimidad de los sujetos de investigación y la confidencialidad de los datos obtenidos, entre los principios que deben regir en cualquier tipo de investigación (Ley 25.467, Art. 3). El glosario de la Res. 753-E/2016 los denomina *datos sensibles*, y los define como aquellos que pueden generar algún daño si se difunden y, por lo tanto, deben mantenerse en reserva. La misma norma menciona los derechos personalísimos que podrían verse vulnerados por la difusión: derecho al honor; derecho a la intimidad; derecho a la imagen (Art. 9.5). También se contempla la confidencialidad de los datos primarios o resultados no publicados ni patentados, y reserva a las instituciones la debida justificación de esa exclusión (Ley 26.899, Art. 6; Res. 753-E/2016, Art. 9.5), esto alcanza la producción obtenida en el marco de acuerdos de confidencialidad con terceros que financiaron la investigación. En concordancia, la Resolución 640-E/2017, establece en su glosario que los datos abiertos no deben estar sujetos a derechos de autor, patentes, marcas comerciales o regulación de secretos comerciales.

El tema de la confidencialidad de los datos toca de cerca disciplinas que, debido a su naturaleza específica, trabajan con datos sensibles y cuyos resultados tienen implicancias directas para la vida de las personas. Es en el marco de estas investigaciones donde el tópico

de la confidencialidad representa un aspecto que los investigadores deben garantizar en los planes de gestión de datos. Se comprende que sean las áreas que más datos sensibles gestionan, como las ciencias de la salud, las que tengan asumidos estos requerimientos (Melero, R.; Hernández-San-Miguel, J., 2014; Pinfield, Cox y Smith, 2014), incluso antes de los mandatos en este campo.

6.4 Los planes de gestión de datos

Resulta necesario retomar la propuesta planteada por Pinfield, Cox y Smith (2014), según la cual un programa de gestión de datos es un instrumento a nivel institucional (nivel macro), por tanto, corresponde diferenciarlo del plan de gestión de datos, aplicable a nivel de un proyecto de investigación (nivel micro).

Las políticas y normativas explican que sea cada vez más frecuente la necesidad de que los investigadores presenten un Plan de Gestión de Datos como requisito para obtener la aprobación de sus proyectos y recibir financiamiento. Un plan de gestión de datos (PGD) cristaliza de manera escrita la forma en que los investigadores tratarán sus datos a lo largo del ciclo de vida, de modo que aseguren su disponibilidad e integridad para que otros investigadores los puedan consultar a largo plazo. Como se expresa en el glosario de la Resolución 753-E/2016):

es un “documento que define e identifica al responsable de los datos primarios que se generarán en un proyecto de investigación, brinda información sobre los tipos y características de los datos, estándares que se utilizarán, cómo serán explotados, compartidos y accesibles para su verificación y reutilización, las medidas que se tomarán para su conservación y preservación, entre otras cuestiones”.

Como resume el material del curso RDMLA (Foundations of RDM, Research Data Management Librarian Academy, 2020), el PGD contempla todas las etapas del ciclo de vida y responderá a las preguntas: ¿qué datos se recolectarán, generarán y usarán?, ¿cómo serán almacenados?, ¿Cómo serán descritos?, ¿Cómo se les dará acceso? También informará sobre formatos, seguridad, preservación, responsabilidades, recursos, cuestiones legales y éticas relacionadas con los datos. Es necesario que un PGD, contemple la tipología y la política de acceso, archivo, distribución, reutilización y preservación (Melero, R.; Hernández-San-Miguel, J., 2014)

La guía para el uso de la Plataforma PGD CONICET sirve para identificar la información que requiere un PGD a modo de ejemplo (CONICET, 2017):

- Título del plan:** título del proyecto asignado.
- Identificación del titular o responsable del proyecto de investigación.**
- Datos para el contacto con el titular.**

- ❑ **Resumen del proyecto.**
- ❑ **Descripción de los datos generados y recolectados.**
- ❑ **Información sobre entidades y personas relacionadas con la generación de los datos.**
- ❑ **Información sobre esquemas o estándares de metadatos para describir los datos.**
- ❑ **Información sobre organización, almacenamiento y resguardo de los datos:** volumen de datos, dónde se almacenarán y copias de seguridad
- ❑ **Información sobre el acceso y restricciones de los datos.**
- ❑ **Información relativa a la protección de datos sensibles:** anonimización, consentimientos por ej.
- ❑ **Información sobre responsabilidades y distribución de las tareas de gestión de datos entre los miembros del grupo de investigación.**

6.5 Repositorios de datos

Son las instituciones que componen el ecosistema de la producción del conocimiento (universidades, financiadores, centros de investigación, etc.), las que deben proveer a los investigadores la infraestructura y los servicios para que depositen y preserven los datos primarios (Melero, R.; Hernández-San-Miguel, J., 2014). Tanto la Ley 26.899 (Art. 1°), como su reglamentación Resolución 753-E/2016 (Cap. II Art. 3°), recogen este principio en el sistema argentino.

Existen diferentes alternativas para el depósito de los datos de investigación. Una lista completa de los repositorios de datos existentes se puede consultar en el Registry of Research Data Repositories (re3data.org). Los investigadores deberán revisar las condiciones establecidas en las políticas mandatorias para determinar la mejor alternativa. Guiarlos en esta elección es parte de los servicios de apoyo de las bibliotecas, que pueden proponer repositorios que cumplan con los criterios exigidos por el organismo financiador.

- ❑ Repositorios institucionales (p. ej. [Dataverse UNR](#); [Harvard Dataverse](#)).
- ❑ Sistemas a nivel nacional (p. ej. [DACyTAR](#))
- ❑ Repositorios recomendados por entidades financiadoras (p. ej. National Science Foundation [NSF data.gov](#); [NIH Genomic Data Sharing Policy](#))
- ❑ Repositorios de datos temáticos (p. ej. [Portal de Datos Genómicos del SNDG](#); [Structural Biology Data Grid](#); [UK Data Archive](#))
- ❑ Plataformas del sector privado. (por ejemplo [Mendeley Data](#))

Recientemente, el Sistema Nacional de Repositorios Digitales (SNRD) de Argentina, lanzó DACyTAr, un portal de datos abiertos de investigación para acceder a los datos primarios depositados en los repositorios del Sistema Nacional, los primeros repositorios cooperantes del portal son INTA Digital (Instituto Nacional de Tecnología Agropecuaria), RepHipUNR (Universidad Nacional de Rosario), Repositorio Digital Universitario (Universidad Nacional de Córdoba), Repositorio Institucional (Servicio Geológico Minero Argentino), RID (Universidad Nacional de Río Negro) y el SEDICI (Universidad Nacional de la Plata). (Sistema Nacional de Repositorios Digitales, 2020).

La necesidad de contar con plataformas que cumplan las exigencias de la gestión de datos, explica la existencia de iniciativas como los Principios TRUST para repositorios digitales (Lin, D. et al., 2020). Estos promueven la confiabilidad de los repositorios digitales en un ecosistema científico donde cobran cada vez mayor relevancia. La denominación de los principios proviene de sus siglas en inglés:

- **Transparencia (Transparency):** se refiere a políticas claras y verificables acerca de los servicios, los datos almacenados y la preservación digital.
- **Responsabilidad (Responsibility):** asegurar la autenticidad e integridad de los datos y claro compromiso con estándares de metadatos y buenas prácticas de curaduría.
- **Enfoque en el usuario (User focus):** para que se cumplan las expectativas de las comunidades de usuarios. Implica proporcionar métricas de uso, mecanismos de descubrimiento efectivos y monitoreo de la comunidad.
- **Sostenibilidad (Sustainability):** asegurar la planificación a largo plazo, el financiamiento y los mecanismos de gobernanza necesarios para la preservación a largo plazo.
- **Tecnología (Technology):** sostener los servicios con estándares y tecnologías adecuadas, asegurando la capacidad de detectar y responder las amenazas de seguridad.

6.6 Comunicación y publicación de los datos

El protagonismo que adquieren los datos en el contexto de la Ciencia Abierta, ha generado productos o iniciativas relacionadas con la comunicación de los mismos. En esta tesina el término “publicar” se utiliza para referir un proceso formal de revisión, edición y publicación en un medio de comunicación especializado. Dos vías de publicación de los datos de investigación parecen más representativas: la publicación de los datos como complemento de un artículo tradicional, y la publicación de los datos en un *data paper* o *artículo de datos*. La Global Biodiversity Information Facility (GBIF), una red global que provee infraestructura para la información sobre biodiversidad, y que promueve el acceso a los datos sobre cualquier tipo de vida en el planeta, define el “*data paper*” como un documento revisado por

pares cuya finalidad es describir datasets y la forma en que se recolectaron o generaron (GBIF.org Artículos de datos). La Resolución 753-E/2016, traduce “*data paper*” como “*documentos de datos*” y agrega que esta publicación puede incluir un análisis de las técnicas que dan cuenta de la validez de los datos.

En base a lo identificado, hay revistas que publican los artículos acompañados con sus datos de investigación como recurso adicional, y revistas centradas en la publicación de data papers. Plataformas de publicación como F1000Research, publican los artículos y sus datos disponibles de manera rápida y con revisión abierta. Las revistas de PLOS ONE requieren a los investigadores que los datos de investigación se encuentren disponibles sin restricciones al momento de publicar, lo mismo plantea la editora de acceso abierto Ubiquity Press. Revistas como Journal of Open Health Data, Data in Brief, Data Science Journal o Earth System Science Data, por nombrar algunas, publican investigaciones sobre datos científicos, su gestión, diseminación o reutilización. Scientific Data, es una revista de la editora Springer Nature cuya finalidad es la publicación de descripciones de datasets, un tipo de publicación que la propia revista denomina “*Data Descriptor*”. Al momento de realizar esta investigación, no se han identificado revistas latinoamericanas con políticas u objetivos similares. Un estudio profundo sobre la publicación de datos primarios y sus beneficios para promover su calidad, citación y disponibilidad fue publicado en la revista Journal of the Association for Information Science and Technology por Candela y otros (Candela, Castelli, Manghi, y Tani, 2015).

Para que un trabajo con estas características sea publicado, debe cumplir criterios similares a los de un artículo tradicional. Algunos de estas condiciones incluyen que los datos cuenten con metadatos adecuados, pasen una revisión de calidad, estar indexados y ser localizables, ser citables en artículos (Bezjak, S. y otros, 2019)

Los datos son vinculados a través de identificadores como el DOI (Identificador de objeto Digital), asignado por servicios como DataCite. La vinculación de los datos con identificadores persistentes que garanticen una identidad única es de gran importancia para su aprovechamiento en los procesos de investigación. Bezjak, S. y otros (2019), señalan la importancia de citar adecuadamente los datos en las revistas y que la vinculación de los artículos con sus datos primarios promoverá nuevas formas de evaluación y publicación.

6.7 Los principios FAIR

En 2014 representantes de la comunidad académica se reunieron en el Centro Lorentz en Leiden para un taller organizado por el Centro eScience de los Países Bajos y el Techcentre holandés para las Ciencias de la Vida (DTL), durante el cual acordaron una serie de buenas prácticas aplicables a datos, herramientas e infraestructuras, para garantizar el descubrimiento y utilización de los datos en el marco de una ciencia que los produce y utiliza de manera masiva ([FORCE11](#), 2016). Estos principios, orientados a mejorar la capacidad de procesamiento automático, definen características y prácticas para que los datos y las herramientas necesarias para utilizarlos puedan ser considerados “FAIR”. Estas buenas

prácticas se conocen como Principios FAIR. Según este acuerdo, los datos científicos abiertos deben ser localizables (*findable*), accesibles (*accessible*), interoperables (*interoperable*) y reutilizables (*reusable*):

- ❑ **Localizables:** humanos y computadoras deben poder encontrar los datos y sus correspondientes metadatos.
- ❑ **Accesibles:** datos y metadatos se deben poder recuperar mediante su identificador utilizando un protocolo de comunicaciones estandarizado y abierto.
- ❑ **Interoperables:** el formato de los datos y de los metadatos debe ser abierto e interpretable por distintos sistemas.
- ❑ **Reutilizables:** la descripción de los datos y los metadatos deben facilitar la reutilización, y deben estar asignadas las licencias que indiquen claramente sus condiciones de uso.

Las decisiones sobre los metadatos y los identificadores persistentes son de vital importancia para cumplir con los principios FAIR y lograr que los datos sean finalmente recuperables y reutilizables. Estos elementos no sólo facilitan el acceso a los humanos, sino que también posibilitan el descubrimiento automatizado (Bezjak, S. y otros, 2019).

La siguiente tabla está basada en la literatura revisada (FORCE11, 2016; Bezjak, S. y otros, 2019; Wilkinson, 2016; GO FAIR, 2017), detalla las condiciones componentes de los principios FAIR:

Principios FAIR			
Principios		Condiciones	
F	Localizables	F.1	Datos y metadatos cuentan con un identificador único y persistente a lo largo del tiempo.
		F.2	Los datos se describen con metadatos enriquecidos que permiten una descripción detallada y relevante (en los términos de R.1)
		F.3	Los metadatos incluyen de manera explícita y clara el identificador del conjunto de datos que describen.
		F.4	Los metadatos y los datos están indexados en una herramienta de búsqueda.
A	Accesibles	A.1	A.1.1. El protocolo de comunicación es abierto, gratuito y universalmente utilizable.

			A.1.2. El protocolo permite implementar un procedimiento de autenticación y autorización cuando fuere necesario.
		A.2	Los metadatos son accesibles aún cuando los datos no estén disponibles.
I	Interoperables	I.1	Datos y metadatos utilizan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento.
		I.2	Datos y metadatos usan vocabularios controlados FAIR.
		I.3	Los metadatos incluyen referencias calificadas (cruzadas), que clarifican las relaciones entre los conjuntos de datos, enriqueciendo el conocimiento contextual de los datos.
R	Reutilizables	R.1	Los metadatos permiten una descripción detallada y relevante. El estándar de metadatos no sólo permite localizar los datos, sino que describe el contexto en que fueron generados.
			R.1.1. los metadatos se liberan con una licencia de uso de datos clara y accesible para humanos y computadoras.
			R.1.2. los metadatos identifican la procedencia de los datos. Esto incluye metadatos sobre su origen, historial y cómo citarlos.
			R.1.3. Los metadatos cumplen con los estándares relevantes para el dominio temático.

La reglamentación de la Ley 26.899 establece que los datos primarios pueden incluir licencias de uso que indiquen ciertas condiciones para los usuarios finales (Res. 753-E/2016). Para que un dato integre el [Portal de Información de Ciencia y Tecnología Argentino](#), debe incluir una licencia abierta (Resolución 640-E/2017, art. 5). Las condiciones que incluyen las licencias de uso, además de reconocer la autoría, también protegen ciertos elementos de la integridad de los datos.

6.8 La cita de los datos

Melero, R.; Hernández-San-Miguel, J. (2014) indican que los conjuntos de datos (datasets) pueden ser citados como cualquier otro producto de la investigación y se les puede asignar una licencia de uso e identificadores únicos. Al respecto, señalan que gracias a la utilización del DOI (Digital Object Identifier), los datasets se pueden relacionar con los autores a través del ORCID (Open Researcher and Contribution ID). DataCite, una organización cuya misión es proveer identificadores persistentes (PIDs) para los datos y otros resultados de

investigación, considera que los datos se pueden citar como cualquier otra fuente de información (Data Cite).

La Declaración Conjunta de Principios de Citas de Datos (Data Citation Synthesis Group, 2014), afirma que los datos son un producto de investigación citable, y reconoce que están en la base de los estudios científicos. Por este motivo, la cita de los datos es una buena práctica de investigación. La Declaración establece los principios a ser adoptados por la comunidad académica para promover además, que la citación de los datos sea comprensible tanto por humanos como por computadoras. Aquí se citan brevemente: 1) **Importancia**: los datos tienen la misma importancia que cualquier otro tipo de publicación y deben ser citados; 2) **Crédito y atribución**: las citas deben facilitar el reconocimiento de los autores y de los créditos académicos; 3) **Evidencia**: los datos deben ser citados como evidencia de una afirmación científica; 4) **Identificación única**: los datos citados deben tener un identificador único, persistente y utilizado por la comunidad académica; 5) **Acceso**: la cita debe facilitar el acceso a los datos, metadatos y documentos asociados; 6) **Persistencia**: identificadores y metadatos deben persistir independientemente de la vida útil de los datos; 7) **Especificidad y verificabilidad**: las citas permiten la identificación correcta, el acceso y la verificación de los datos. 8) **Interoperabilidad y flexibilidad**: los métodos de citas deben ser flexibles para responder a las necesidades de las comunidades académicas, pero también deben ser interoperables entre esas comunidades.

6.9 La formación y actualización profesional

Parece evidente que el trabajo con datos de investigación está dibujando un perfil especializado dentro de la bibliotecología que requiere una formación especial y una estrategia de actualización para los graduados. Esta necesidad emergente respecto de los planes de formación en las carreras de bibliotecología, es abordada por diferentes autores. Un estudio realizado por Harris-Pierce y Yan Quan Liu en 2012, señalaba que sólo el 31% de los programas en Estados Unidos incluía el área de gestión de datos (Antell y otros, 2014). Es interesante señalar que muchos bibliotecarios acceden a cursos y programas de formación como parte de sus planes de formación continua, o apoyados por las bibliotecas donde se desempeñan (Antell y otros, 2014). Brochu y Burns (2019) señalan que es necesario poner a disposición de las bibliotecas más capacitación en ciencia de datos para mejorar las habilidades de los bibliotecarios y concluyen que la información sobre el rol del bibliotecario en este campo se está incrementando no sólo con artículos, sino con recursos para entrenamiento. Un trabajo realizado por Barbrow, Brush y Goldman (2017) afirma que es necesario ofrecer capacitación y entrenamiento para los bibliotecarios, y propone diversos recursos de formación:

- Listas de discusión (dss-dil_dg@lists.ala.org, dss-rdm_dg@lists.ala.org, CANLIB-DATA Listserv);

- Plataformas digitales colaborativas y sitios especializados ([ResearchDataQ](#), [NNLM RD3](#));
- Materiales generados en talleres, simposios y conferencias (NER e-Science Symposium, Midwest Data Librarian Symposium, [talleres](#) realizados por las bibliotecas del MIT);
- Cursos de entrenamiento en línea ([DataONE Modules](#), [Data Management Skill building Hub](#), [New England Collaborative Data Management Curriculum](#), Purdue Data Curation Online Toolkit);
- Guías temáticas ([Digital Humanities Data Curation Guide](#));
- Recursos multimedia, blogs y redes sociales ([Data ab Initio](#)).

Una experiencia interesante que se puede destacar en este sentido, es el curso Research Data Management Librarian Academy (Research Data Management Librarian Academy, 2020) organizado por grandes bibliotecas universitarias norteamericanas junto a la editora internacional Elsevier, y en el que este autor ha participado. También existen programas formales dentro de las universidades a nivel de posgrado ([University of North Texas Digital Curation and Data Management Graduate Academic Certificate](#)). Nitecki y Davis (2019) destacan las propuestas desarrolladas por la Association of College and Research Libraries de la ALA (ACRL), que incluyen la creación de la sección especial [Digital Scholarship Section](#), con varios grupos de interés relacionados con datos de investigación.

7. CONCLUSIÓN

Por lo expuesto hasta aquí se puede afirmar que el sistema de investigación y desarrollo está atravesado por nuevas herramientas computacionales que posibilitan un ecosistema de datos masivos. Los investigadores generan y utilizan una amplia gama de datos que luego analizan para crear nueva información. El valor de esta información, tanto para el presente como para el futuro, está reconocido en diferentes mandatos institucionales y políticas que exigen distintas acciones para garantizar la integridad y disponibilidad de los datos. Todas estas medidas están imbricadas a lo largo de un ciclo de vida que requiere la colaboración entre los investigadores y diversos actores dentro de las instituciones, destacándose el rol de las bibliotecas. Tanto las políticas sobre gestión de datos como las experiencias desarrolladas por múltiples bibliotecas universitarias de diversos países, demuestran que los bibliotecarios pueden y deben participar en la gestión de los datos de investigación, brindando servicios especializados a sus comunidades. Se observa una creciente cristalización de ese rol y el consiguiente reconocimiento por parte de los distintos interesados, algo que Pinfield, Cox y Smith (2014), señalaban como “*factor jurisdiccional*” y que se manifiesta en políticas y declaraciones. Witt considera que los datos de investigación serán “*un componente maduro de la bibliotecología cuando sea aceptada como una práctica regular cuando términos como ‘referencia de datos’, se conviertan en simple ‘referencia’, y los conjuntos de datos no*

reciban más tratamiento específico o especializado que cualquier otra colección” (2012 p. 186)

No obstante, se debe atender que, si bien las experiencias (fundamentalmente en el ámbito anglosajón y europeo) indican una relación cierta de las bibliotecas académicas en el proceso de gestión de los datos, las iniciativas no son uniformes. La baja incidencia de estudios parece indicar que las bibliotecas de América Latina no se encuentran todavía, en el mismo nivel de implicación. La consolidación de una tendencia en el campo profesional está estrechamente ligada a la generación de políticas a nivel gubernamental e institucional.

Las bibliotecas se encuentran ante la posibilidad de una expansión en términos de servicios. Los estudios muestran que se está consolidando un nuevo rol de los bibliotecarios a partir, no solo del activo informativo, sino de un nuevo lugar de interacción con distintos actores de la comunidad. Los bibliotecarios parten de una plataforma sólida para asumir las tareas que requiere este nuevo rol. Este punto de partida favorable está formado, al menos, por dos factores importantes: el primero es la misma existencia de la biblioteca, una estructura tradicional y estable, que constituye un marco reconocido y fiable para todos los actores de la comunidad; el segundo factor es el conjunto de habilidades y competencias tradicionales de los bibliotecarios, para quienes el tratamiento de la información para su acceso, está en el núcleo de su actividad. Estos dos factores son potenciadores para la reconfiguración hacia el bibliotecario de datos, lo cual no debe subestimar las complejidades de este campo de acción, ni restar importancia a la necesidad de incorporar nuevos conocimientos y entrenar nuevas habilidades.

Sin quitar mérito a la oferta de cursos y espacios de formación emergentes como los MOOCs (cursos virtuales masivos), será fundamental que los programas de formación académica en Bibliotecología y Documentación promuevan las competencias necesarias para formar los perfiles profesionales adecuados. Este trabajo es un aporte para pensar un programa mínimo de formación introductoria a la materia.

Otra situación emergente que resulta de especial importancia y amerita mayores reflexiones, es la naturaleza interdisciplinar del proceso de gestión de los datos y la necesidad de promover marcos de cooperación entre los bibliotecarios y las distintas áreas dentro de las instituciones. Los bibliotecarios, siendo parte de esos equipos de trabajo, deben desarrollar una *conciencia* sobre su propia participación, de modo tal que la reflexión sobre la praxis contribuya a una participación activa. En el sentido de una interdisciplina proactiva como propone Tripaldi (2018), los bibliotecarios deben participar en la creación de las metodologías y en la definición de los objetivos que se fijan los equipos de trabajo. Todo beneficio de la cooperación debe resultar en la mejor gestión de los activos digitales y en el fortalecimiento del rol institucional de los actores implicados.

La cuestión de los recursos que estos servicios plantean, es un factor a monitorear. Que las bibliotecas de la región se sumen a la provisión de servicios de datos requiere desarrollar infraestructuras, pero también la formación y contratación de personal adecuado.

En relación a la disponibilidad de investigación sobre el tema, resulta interesante rescatar lo expuesto por Brochu y Burns (2019), que remarcan la necesidad de incrementar la indagación desde el campo bibliotecario en la forma de estudios de casos, ya que este tipo de publicación provee experiencias y ejemplos prácticos reales.

8. BIBLIOGRAFÍA

- Anglada, L. M., & Abadal, E. (2018). ¿Qué es la ciencia abierta? Anuario ThinkEPI, 12, 292-298. <https://doi.org/https://doi.org/10.3145/thinkepi.2018.43>
- Argentina. Jefatura de Gabinete de Ministros. Resolución 538/2013 (2013). Argentina. Recuperado a partir de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/215000-219999/218131/norma.htm>
- Argentina. Ley 25.467 Ciencia, Tecnología e Innovación. (2001). Argentina. Recuperado a partir de <https://www.argentina.gob.ar/normativa/nacional/ley-25467-69045/actualizacion>
- Argentina. Ley 26.899 Repositorios digitales institucionales de acceso abierto. (2013). Argentina. Recuperado a partir de http://repositorios.mincyt.gob.ar/pdfs/Boletin_Oficial_Ley_26899.pdf
- Argentina. Ministerio de Ciencia Tecnología e Innovación Productiva. Resolución 753 - E/2016 (2016). Argentina. Recuperado a partir de <http://www.mincyt.gob.ar/financiamiento>
- Argentina. Ley 27.275 Derecho a la información pública. (2016). Recuperado a partir de <https://www.argentina.gob.ar/normativa/nacional/ley-27275-265949/texto>
- Argentina. Ministerio de Ciencia, Tecnología e Innovación. (2020). Portal de Información de Ciencia y Tecnología Argentino. Recuperado 3 de noviembre de 2020, a partir de <https://datos.mincyt.gob.ar/#/>
- Argentina. Ministerio de Ciencia Tecnología e Innovación Productiva. Resolución 640 - E/2017 (2017). Argentina. Recuperado a partir de <https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-640-2017-277216>
- Argentina. Ministerio de Modernización. Decreto 117/2016 (2016). Argentina. Recuperado a partir de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/257755/norma.htm>
- Argentina. Ministerio de Modernización. Res. 11/2016 (2016). Argentina. Recuperado a partir de <https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-11-2016-259326/texto>

- Argentina. Poder Ejecutivo Nacional. Decreto 1172/2013 (2003). Argentina. Recuperado a partir de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/90000-94999/90763/norma.htm>
- Argentina. Poder Ejecutivo Nacional. Decreto 206/2017 (2017). Argentina. Recuperado a partir de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/270000-274999/273023/norma.htm>
- Argentina. Sistema Nacional de Repositorios Digitales. (2020). DACyTAr. Recuperado 30 de octubre de 2020, a partir de <https://dacytar.mincyt.gob.ar/>
- Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries*, 75(4), 557-574. <https://doi.org/10.5860/crl.75.4.557>
- Barbrow, S., Brush, D., & Goldman, J. (2017). Research data management and services Resources for novice data librarians. *C&RL News*, 274-278. <https://doi.org/https://doi.org/10.5860/crln.78.5.274>
- Bezjak, S., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., Kramer, B., ... Heller, L. (2019). Manual de capacitación sobre Ciencia Abierta. Léeme · GitBook. Foster Open Science. <https://doi.org/https://doi.org/10.5281/zenodo.1212496>
- Brochu, L., & Burns, J. (2019). Librarians and Research Data Management—A Literature Review: Commentary from a Senior Professional and a New Professional Librarian. *New Review of Academic Librarianship*. <https://doi.org/10.1080/13614533.2018.1501715>
- Brown, E. (2010). 'I Know What You Researched Last Summer': How Academic Librarians Are Supporting Researchers in the Management of Data Curation. *The New Zealand Library & Information Management Journal*, 52(1), 55-69. Recuperado a partir de https://lianza.org.nz/sites/default/files/files/nzlimj_vol_52_issue_no_1_oct_2010.pdf
- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747-1762. <https://doi.org/10.1002/asi.23358>
- Carlson, J. (2012). Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data. *Reference Services Review*, 40(1), 7-23. <https://doi.org/10.1108/00907321211203603>
- Choudhury, G. S. (2008). Case study in data curation at John Hopkins University. *Library Trends*, 57(2), 211-220. Recuperado a partir de

<https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/34023/57.2.choudhury.pdf>

Choudhury, G. S. (2008). Case study in data curation at John Hopkins University. *Library Trends*, 57(2), 211-220. Recuperado a partir de <https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/34023/57.2.choudhury.pdf>

cOAlition S. (2018). Plan S: Principles and Implementation. Recuperado 11 de junio de 2020, a partir de <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>

CODATA. (2020). *Data Science Journal*. Recuperado 3 de noviembre de 2020, a partir de <https://datascience.codata.org/>

Consejo Nacional de Investigaciones Científicas y Técnicas. (2017). Guía usuario: confección de un Plan de Gestión de Datos PGD CONICET. Recuperado a partir de <http://pgd.conicet.gov.ar/files/guia-usuario-pgd-conicet.pdf>

Copernicus Publications. (2020). *Earth System Science Data*. Recuperado 27 de junio de 2020, a partir de <https://www.earth-system-science-data.net/>

Data Citation Synthesis Group. (2014). Joint Declaration of Data Citation Principles. FORCE11. Recuperado a partir de <https://doi.org/10.25490/a97f-egykh>

DataCite International Data Citation Initiative. (2020). *DataCite*. Recuperado 27 de junio de 2020, a partir de <https://datacite.org/>

Data Observation Network for Earth. (s. f.). *Data Life Cycle DataONE*. Recuperado 24 de julio de 2020, a partir de <https://old.dataone.org/data-life-cycle>

Declaración de San Francisco sobre la evaluación de la investigación. (2018). *Revista ORL*, 9(4), 295-299. Recuperado a partir de <https://doi.org/10.14201/orl.17845>

Dietrich, D., Adamus, T., Miner, A. and Steinhart, G. (2012) De-Mystifying the Data Management Requirements of Research Funders,” *Issues in Science and Technology Librarianship* 70. Recuperado a partir de www.istl.org/12-summer/refereed1.html

Digital Curation Centre. (2018). An analysis of open data and open science policies in Europe, v2.1. Recuperado a partir de <https://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/>

Digital Curation Centre. (2020). An analysis of open data and open science policies in Europe, v6. Recuperado a partir de <https://zenodo.org/record/4005612#.X5x94G5KjIU>

- Elsevier. (2020). Data in Brief. Recuperado 27 de junio de 2020, a partir de <https://www.journals.elsevier.com/data-in-brief>
- F1000 Research Ltd. (2020). F1000Research. Recuperado 27 de junio de 2020, a partir de <https://f1000research.com/>
- Faniel, I. M., & Connaway, L. S. (2018). Librarians' Perspectives on the Factors Influencing Research Data Management Programs. *College & Research Libraries*, 79(1), 100-119. <https://doi.org/https://doi.org/10.5860/crl.79.1.100>
- FORCE11. (2016). The FAIR Data Principles. Recuperado 26 de julio de 2020, a partir de <https://www.force11.org/group/fairgroup/fairprinciples>
- Garritano, J. R., & Carlson, J. R. (2009). A subject librarian's guide to collaborating on e-science projects. *Issues in Science and Technology Librarianship*, 57. Recuperado a partir de <http://www.istl.org/09-spring/refereed2.html#15>
- Global Biodiversity Information Facility. (s. f.). Artículos de datos. Recuperado 27 de junio de 2020, a partir de <https://www.gbif.org/es/data-papers>
- Go FAIR. (2017). FAIR Principles. Recuperado 26 de julio de 2020, a partir de <https://www.go-fair.org/fair-principles/>
- Heidorn, P. B., (2011) The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, 51(7/8), 662–72. Recuperado a partir de <https://www.tandfonline.com/doi/abs/10.1080/01930826.2011.601269>
- Higgins, S. (2007). Draft DCC Curation Lifecycle Model Sarah. *Proceedings of the ACM International Conference on Digital Libraries*, 2(2), 453. <https://doi.org/10.1145/1378889.1378998>
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134-140. <https://doi.org/10.2218/ijdc.v3i1.48>
- Hswe, P. y Holt, A. (2011) Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, 274, 11–17. Recuperado a partir de <https://doi.org/10.29242/rli.274.2>.
- Jsic. (s. f.). Juliet Statistics - v2.sherpa. Recuperado 27 de junio de 2020, a partir de https://v2.sherpa.ac.uk/view/funder_visualisations/1.html
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., ... Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data. Nature Research*. <https://doi.org/10.1038/s41597-020-0486-7>

- Luce, R. (2008). Learning from E-Databases in an E-Data World. *Educause Review*, 43(1), 12-13. Recuperado a partir de <https://er.educause.edu/articles/2008/1/learning-from-edatabases-in-an-edata-world>
- Matusiak, K. K., & Sposito, F. A. (2017). Types of research data management services: An international perspective. *Proceedings of the Association for Information Science and Technology*, 54(1), 754-756. <https://doi.org/10.1002/pr2.2017.14505401144>
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5(JULY), 1-19. <https://doi.org/10.7554/eLife.16800>
- Melero, R.; Hernández-San-Miguel, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española de Documentación Científica*, 37(4): e066. doi: <http://dx.doi.org/10.3989/redc.2014.4.1154>
- Mering, M. (2020). Open Access Mandates and Policies: The Basics. *Serials Review*, 1-3. Recuperado a partir de <https://doi.org/10.1080/00987913.2020.1760707>
- Mullins, J. L. (2007). Enabling International Access to Scientific Data Sets : Creation of the Distributed Data Curation Center (D2C2). *Libraries Research Publications*. Recuperado a partir de https://docs.lib.purdue.edu/lib_research/85/
- National Institutes of Health. (2019). Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance. Recuperado a partir de <https://www.federalregister.gov/documents/2019/11/08/2019-24529/request-for-public-comments-on-a-draft-nih-policy-for-data-management-and-sharing-and-supplemental>
- Nitecki, D. A., & Davis, M. E. K. (2019). Expanding academic librarians' roles in the research life cycle. *Libri*, 69(2), 117-125. <https://doi.org/10.1515/libri-2018-0066>
- Olaz, Á. (2011). Una aproximación conceptual a la cualificación profesional desde una perspectiva competencial. *Papers*, 96(2), 589-616. Recuperado a partir de <https://papers.uab.cat/ojs-papers/papers/article/view/v96-n2-olaz/373>
- Olaz Capitán, A. J., Ortiz García, P., & Sánchez-Mora Molina, M. I. (2011). Una aproximación conceptual al término competencia desde un análisis polisémico. En *Universidad Politécnica de Cartagena (Ed.), Congreso Internacional de Innovación Docente* (pp. 129-138). Cartagena. Recuperado a partir de <http://repositorio.upct.es/bitstream/handle/10317/2079/c13.pdf?sequence=1>

- Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLoS ONE*, 9(12), 1-28. <https://doi.org/10.1371/journal.pone.0114734>
- PLOS ONE. (2019). PLOS ONE: data availability. Recuperado 27 de junio de 2020, a partir de <https://journals.plos.org/plosone/s/data-availability>
- re3data.org Project Consortium. (2020). re3data.org Registry of Research Data Repositories. Recuperado 3 de noviembre de 2020, a partir de <https://doi.org/10.17616/R3D>
- Read, K. B., Koos, J., Miller, R. S., Miller, C. F., Phillips, G. A., Scheinfeld, L., & Surkis, A. (2019). A model for initiating research data management services at academic libraries. *Journal of the Medical Library Association*, 107(3), 432-441. <https://doi.org/10.5195/jmla.2019.545>
- Research Data Management Librarian Academy Partnership. (2020). Research Data Management Librarian Academy. Recuperado a partir de <https://rdmla.github.io/>
- ROARMAP. (s. f.). Registry of Open Access Repository Mandates and Policies (ROARMAP). Recuperado 7 de junio de 2020, a partir de <http://roarmap.eprints.org/>
- Schmidt, B. & Shearer, K. (2016). Librarians' Competencies Profile for Research Data Management. En: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Proceedings of the 20th International Conference on Electronic Publishing. Recuperado a partir de https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf
- Tejada, C., & Rodríguez, L. (2003). Sistematizaciones de competencias de los profesionales de la información. Valoración de la relación de Decidoc por los asociados de Sedic. *El Profesional de la Información*, 12(1), 10-17. Recuperado a partir de <http://elprofesionaldeinformacion.metapress.com/media/h82chcayxjduygfpty47/contributions/p/6/9/d/p69d82w0vcv86qbe.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. Chicago, IL. Recuperado a partir de http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf
- Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., Lundeen, A. (2015). Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship*, 4(2), e1085. <https://doi.org/10.7191/jeslib.2015.1085>

- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., ... Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER QUARTERLY*, 27(1), 23-44. <https://doi.org/10.18352/lq.10180>
- Tripaldi, N. M. (2018). Ecos epistemológicos de la reglamentación de la Ley 26.899 (Repositorios Digitales Institucionales de Acceso Abierto): Interdisciplina y Bibliotecología. En 20 Jornadas de Bibliotecas Agropecuarias, Veterinarias, Forestales y Pesqueras, 50 Reunión Nacional de Bibliotecarios. Recuperado a partir de <http://hdl.handle.net/20.500.12123/2301>
- Ubiquity Press. (s. f.). *Journal of Open Health Data*. Recuperado 27 de junio de 2020, a partir de <https://openhealthdata.metajnl.com/about/>
- Ubiquity Press. (2020). *Research integrity*. Recuperado 27 de junio de 2020, a partir de <https://www.ubiquitypress.com/site/research-integrity/>
- University of California Curation Center. (2020). *DMPTool v3.0.0*. Recuperado 1 de noviembre de 2020, a partir de https://dmptool.org/general_guidance#types-of-data
- Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing Data Curation Profiles. *International Journal of Digital Curation*, 4(3), 93-103. <https://doi.org/10.2218/ijdc.v4i3.117>
- Witt, M. (2012). Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. *Journal of Library Administration*, 52(2), 172-188. <https://doi.org/10.1080/01930826.2012.655607>

El autor reconoce la invaluable información de las siguientes guías temáticas elaboradas por bibliotecas universitarias:

Biblioteca de la Comisión Económica para América Latina y el Caribe: Gestión de datos de investigación. <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/tipos-datos> consultado 01/11/2020

Biblioteca de la Universitat Oberta de Catalunya: ¿Cómo gestionar y publicar los datos de investigación? <http://biblioteca.uoc.edu/es/investigacion/como-gestionar-y-publicar-los-datos-de-investigacion#apartat-8540> consultado 27/06/2020

Association of College and Research Libraries: Scholarly Communication Toolkit: Research Data Management. <https://acrl.libguides.com/scholcomm/toolkit/datamanagement> consultado 12/10/2020