

Los repositorios institucionales de datos para la investigación y la gestión de la información

MINERVA CASTRO ESCAMILLA

Universidad Nacional Autónoma de México, México

INTRODUCCIÓN

El aumento de la actividad científica produce cambios profundos a nivel mundial con sofisticados avances en Tecnologías de Información y Comunicación (TIC) y en el incremento de la inversión, tanto pública como privada, el avance en áreas como las ciencias de la tierra, el mar y la atmósfera, así como en la astronomía, la física o las ciencias biomédicas y de la vida, está permitiendo la obtención y organización de datos, que, además, crecen de manera exponencial. Como consecuencia, tanto especialistas como organizaciones dedicadas a la investigación se encuentran cada vez con más dificultad para gestionar, analizar e interpretar esta producción de datos en grandes volúmenes.

Hoy en día, las universidades o institutos de investigación son los que, en general, gestionan repositorios, convirtiéndolos en una herramienta en sus políticas científicas y académicas, además de ser una pieza de apoyo para la enseñanza y la investigación cumpliendo con una serie de pautas, políticas y lineamientos de orden internacional.

Desde un punto de vista más conceptual, un repositorio institucional se convierte en un sistema de gestión de contenidos, que ofrece a la comunidad académica un conjunto de servicios para la creación de productos académicos.

Hace un poco más de dos décadas se han establecido varias plataformas de repositorio, cada una con su propio conjunto de beneficios y criterios técnicos, con el firme propósito de gestionar tanto la información como los datos bajo tres líneas de acción: la recopilación, la preservación y la difusión de los contenidos tanto de publicaciones o los datos de los colegios, universidades e instituciones de investigación. Asociado a ello trabajar en colaboración con grupos de investigación, reestructurando una nueva manera de investigar.

En esta época aparecen nuevos servicios en las bibliotecas, así como formatos y tipos de información, como los sets de datos o los llamados *dataset*, el *software* o los resultados de simulaciones, modelos de sistemas, enlaces de objetos con datos, que deben ser almacenados, preservados y difundidos para la generación de nuevo conocimiento.

La ciencia avanza a pasos agigantados y genera nuevas investigaciones cuyos resultados deben ser publicados y ofrecidos a una comunidad científica, que pueda apoyar la multidisciplinariedad para, con ello, se obtenga nuevos datos ya procesados. Es importante mencionar, que, si el investigador no tiene la capacidad de acceder a una publicación científica o a un conjunto de datos, puede estar perdiendo información valiosa para su trabajo diario.

La publicación y difusión de los resultados de las investigaciones es, por tanto, una necesidad obligada, pero que debe hacerse mediante lineamientos, condicionantes y licencias.

Cuando se hace referencia a los repositorios, hay que tomar en cuenta a aquellos que albergan documentos científicos y académicos, en su mayoría. Como parte de su evolución al rápido y cambiante avance de la investigación científica surge la necesidad de desarrollar repositorios de datos, que alojan y preservan datos científicos generados durante el proceso de investigación y nacen en las diversas áreas de la ciencia.

Se han estructurado diversos proyectos internacionales que apoyan el acceso abierto a la investigación. Los proyectos como DRIVER de la Unión Europea en la Digital Repository Infrastructure Vision for European Research (Ethniko, 2009, p. 17); PIRUS: Publisher and Institutional Repository Usage Statistics (IRUSK-UK, 2009, p. 6) para relacionar estadísticas; EuroCRIS para unificar sistemas de gestión de la investigación y repositorios, los cuales han servido para mejorar la cooperación desde nivel local, regional, nacional e internacional, con la finalidad de ayudar y reforzar a las estrategias y políticas.

En México, en el 2014, se creó una iniciativa para la integración del capítulo 10 de ciencia y tecnología sobre acceso abierto a la información científica y tecnológica, creando el *Programa de Repositorios del Consejo Nacional de Ciencia y Tecnología* (Conacyt) y, a su vez, implementándose el *Repositorio Nacional* (RN). A finales de ese mismo año se da pie a crear los *Lineamientos Generales para el Repositorio Nacional* (RN) y los *Repositorios Institucionales* (RI). Proyectos que han servido para mejorar la cooperación desde nivel local, regional, nacional e internacional, con la finalidad de ayudar a reforzar a las estrategias y el desarrollo de la investigación.

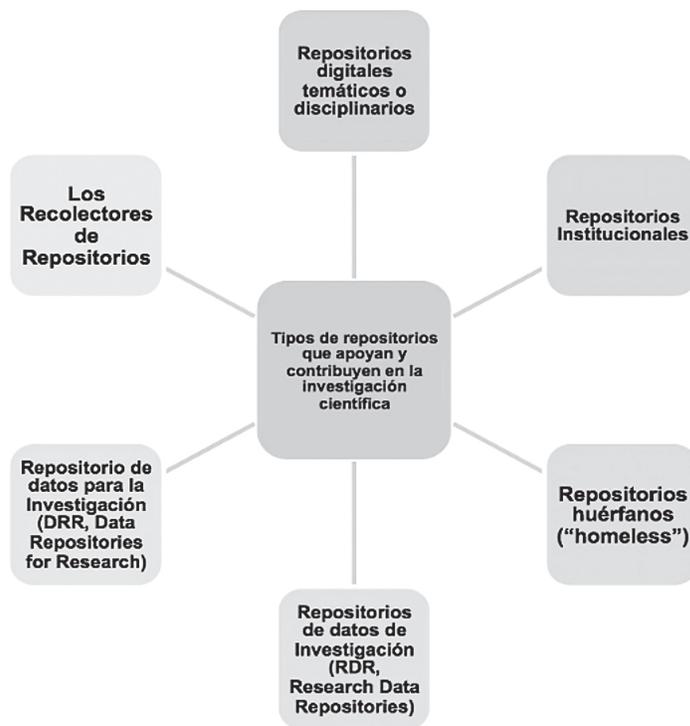
Los repositorios institucionales se están convirtiendo, por ende, en productos de información, que permiten la gestión de contenidos y datos sobre las diferentes disciplinas de especialización a nivel mundial.

LA TIPOLOGÍA DE LOS REPOSITARIOS

En el contexto de la ciencia abierta, y con base en el desarrollo que tiene esta nueva infraestructura de servicios se ha generado una serie de repositorios con diferentes objetivos, que se transforman en herramientas por las cuales se recupera información y datos científicos disponibles, en su mayoría, en acceso abierto o, en algunos casos, con restricciones por las características de la literatura incorporada. Después de analizar y revisar documentación,

que refiere a tipos de repositorios, entre ellas se encuentra una muy destacada realizada por Acosta *et al.* (2018, p. 343), complementando con una investigación y revisión en la literatura dentro del contexto de la ciencia abierta, se consideró la tipología, que se muestra en la Figura 1, como una constante en el área académica y científica.

Figura 1. Tipología de los Repositorios¹



Fuente: Desarrollo de un repositorio de datos científicos de apoyo a la investigación: el caso de las ciencias de la tierra (Castro, 2020, 18).

1 La figura presenta la tipología de los Repositorios dentro del ámbito de la ciencia abierta.

Cada uno de estos debe establecer los parámetros en torno a los tipos de contenido que se alojará y dependerá, en gran medida, del modelo interesado en desarrollar.

Hoy en día, las universidades o institutos de investigación son los que, en general, gestionan este tipo de repositorios, constituyendo una herramienta clave de sus políticas científicas y académicas, además de una pieza de apoyo para la enseñanza y la investigación; por lo que no es correcto llamarle base de datos, pues deben cumplir una serie de pautas, políticas, lineamientos, pero sobre todo apegarse a lo que implica el acceso y la ciencia abierta. Es importante comentar que los documentos albergados, particularmente requieren de los estándares y de una mayor flexibilidad por parte de las plataformas de *software*.

LAS DIRECTRICES, LINEAMIENTOS Y PRINCIPIOS

Para fortalecer la ciencia y su compartición, así como el reúso de datos por parte de los investigadores a nivel mundial, surgen iniciativas desde 2002 a escala mundial con lineamientos adaptables a cada país dependiendo de sus circunstancias y necesidades, sin salirse de la normatividad establecida y, sobre todo, sin dañar a terceros.

En ese sentido, para que un proyecto de Ciencia Abierta fructifique, la base debe contar con tres elementos clave: el libre acceso, que sea apoyada con recursos públicos y facilitar la consulta de información a todos los niveles por medio de herramientas. Es entendible que las normativas internacionales sean lo más completas posibles, pero también es permisible un trabajo cooperativo de distintos tipos y multidisciplinario, que atienda necesidades locales en beneficio de la sociedad; de tal manera, que puede observarse en el análisis de las declaraciones, que cada una toca lineamientos generales y específicos según lo que busca promover.

En el caso de México, contar con una Ley de Ciencia y Tecnología, que integra al acceso abierto y a la información científica, así como los Lineamientos Generales de Ciencia Abierta de Conacyt

(Conacyt, 2017a, p. 3), coloca en una situación comprometida de crecimiento a todos los especialistas, obligándolos a generar proyectos del alcance del Repositorio Nacional, el cual tiene como objetivo abrir sus recursos a la sociedad e intercambiar información para la generación de nueva ciencia; participando, así, en la iniciativa de Ciencia Abierta y cosecha de información de los diferentes repositorios institucionales.

Con todas las manifestaciones de acceso libre, es la sociedad quien se beneficia, ya que existe una igualdad en su consulta con el establecimiento de distintos mecanismos, que permitan su adaptación de manera local y presenten los lineamientos que sirvan de guía entre los grupos académicos.

Todas las iniciativas contribuyen a mejorar la metodología y calidad de los procesos y a la gestión de la información, además de apoyar el crecimiento de la adopción del acceso abierto para la generación de nuevo conocimiento. Es imprescindible generar herramientas, como los RI, que permitan una libre consulta.

Debe entenderse que, bajo una crisis mundial, como la que se está viviendo, está en riesgo el acceso a la información de manera libre por la situación económica. No puede prescindirse, que este movimiento logra mayor disponibilidad y accesibilidad de los resultados de las investigaciones científicas financiadas con fondos públicos, lo que amplía las posibilidades de procesos de revisión por pares, acorta los tiempos de publicación y aumenta la transparencia en los trabajos científicos, lo cual logra un alto impacto de la investigación.

LA GESTIÓN DE LA INFORMACIÓN

La necesidad de tener a la mano una adecuada infraestructura de comunicaciones que permita obtener los datos requeridos es hoy en día tema de impacto. Los repositorios juegan un papel importante a medida que surgen necesidades de almacenamiento con el tiempo; es decir, la enorme acumulación de datos, por lo que configurar dichos repositorios facilita la buena ejecución de las

operaciones. Estas herramientas posibilitan tanto depósito como la difusión, lo que permite conocer la producción científica de la institución y difundirla entre los investigadores y la comunidad interesada. Los principios para llevar a cabo una gestión de la información a través de las herramientas efectivas, proporciona una solución integrada para administrar, procesar y proteger tanto los datos como la información (contenidos digitales) almacenando y simplifica la información de búsqueda y recuperación (Lucio Pilla-sagua, 2022, p. 296).

Es importante considerar, que el proceso de la gestión de los datos para la investigación cuenta con diferentes etapas antes de ser incluidos en un RI; por tanto, el trabajo debe regirse bajo normas internacionales llevadas por organismos regentes, con lo cual se ofrecerán elementos competitivos para el flujo y migración de la información y puedan ser localizables, accesibles, interoperables y reutilizados.

Gracias a la velocidad que han proporcionado las nuevas comunicaciones a la información, el alcance para cada uno de los usuarios que la requiera, se hace de forma rápida y eficiente; por ende, las bibliotecas, entidades académicas, instituciones, de cualquier tamaño y complejidad, que quieran subsistir deberán adaptarse a los cambios.

LOS ELEMENTOS EN LA GESTIÓN DE LA INFORMACIÓN

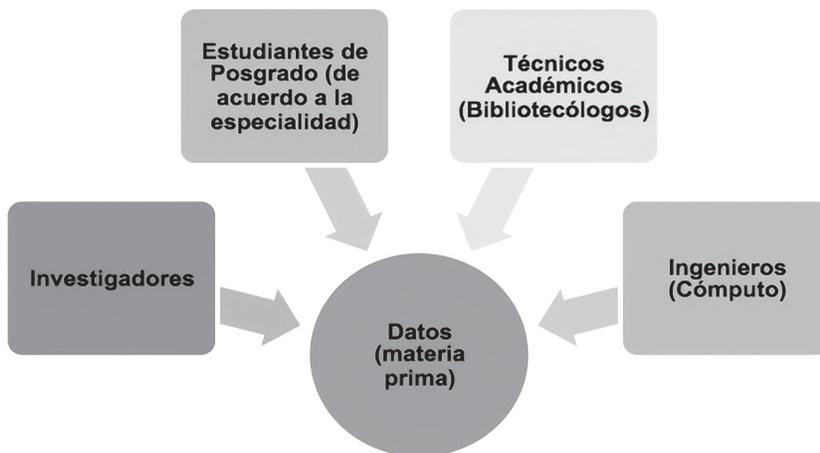
Al referirnos a los elementos de la gestión de la información deben considerarse tanto a nivel interno como externo, así como realizar un análisis completo y detallado del tipo de información; por lo que, es necesario estructurar un grupo de trabajo, que realice dicha actividad y para lograr resultados importantes debe conformarse de la siguiente forma:

- Especialistas de los datos, investigadores o académicos responsables de estos, ya sean generados de forma directa o con la administración de tecnologías relacionadas.

Investigación, reflexión, acción...

- Personal académico y profesionales de la información, quienes manejan la estructura técnica de los metadatos y se plasme, eficientemente, la descripción física y su contexto de cada dato o información, así como sus formas de recuperación.
- Personal de cómputo, parte importante en la instalación de servidores, *software* y configuración de las plataformas y despliegue de información, con esto se logra ejecutar una estrategia de trabajo, que permita un crecimiento constante del proyecto, como puede verse en la Figura 2.

Figura 2. Elementos que participan en la gestión de la información



Fuente: Desarrollo de un repositorio de datos científicos de apoyo a la investigación: el caso de las ciencias de la tierra (Castro, 2020, p. 110).

Como puede verse, se logra un funcionamiento óptimo, además de establecer diferentes mecanismos de control y evaluación, que retroalimentan a los objetivos y a las políticas.

LOS TIPOS DE DATOS EN LOS REPOSITARIOS INSTITUCIONALES

La disponibilidad de flujos de trabajo para la publicación de datos tiene un impacto muy sensible en los científicos, así como en sus prácticas de investigación y en las estrategias de financiación, debido a que muchas veces desconocen cómo dar a conocer lo que generan por medio de plataformas o consideran que los datos son de su propiedad. Actualmente, se requiere que el acceso a los datos sea el resultado de una investigación o publicación de un artículo y también necesita ponerse a disponibilidad de los especialistas los datos duros y puros, para que sirvan como nuevo conocimiento en otras líneas de investigación, que se encuentren disponibles u hospedados en los repositorios correspondientes a su disciplina, fuente de financiación u otros dominios.

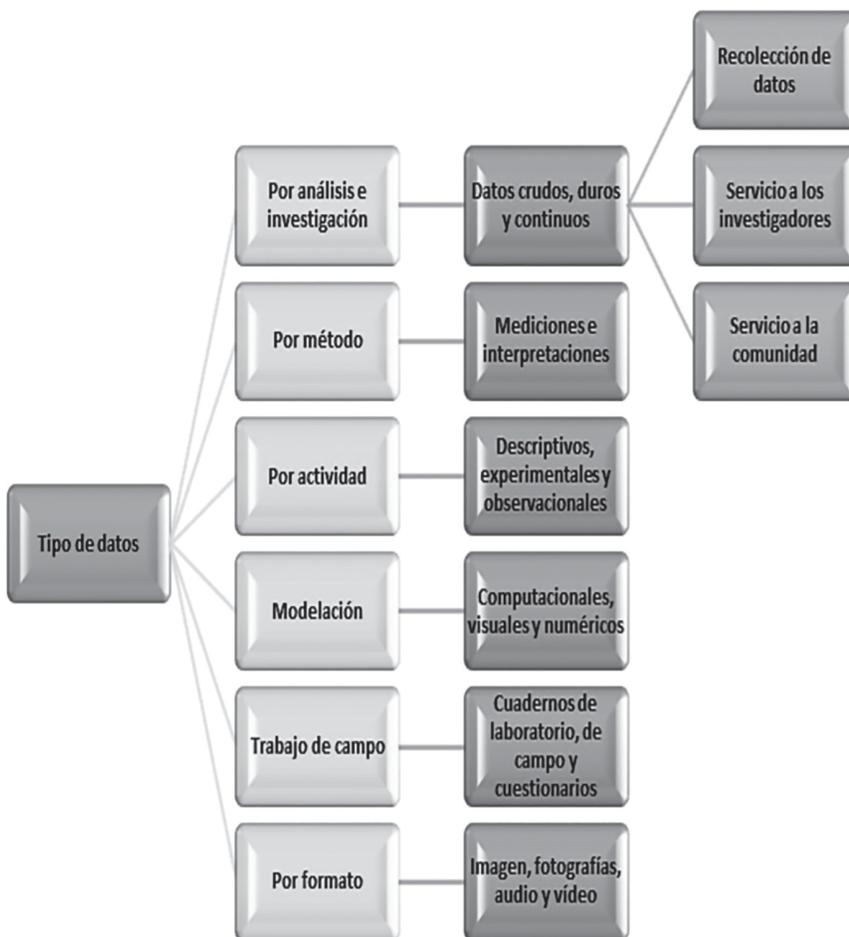
Podemos encontrar datos que se utilizan como fuentes primarias para soportar una investigación técnica o científica, de beca o una actividad artística que, por lo regular, se utiliza como evidencia de un proceso y son los datos más comúnmente aceptados. Podemos partir, incluso, de otra premisa en donde los datos nacen a cada momento y no se interrumpen, aún no pasan por un análisis, uso o interpretación.

Es importante considerar que los contenidos digitales y no digitales tienen el potencial de convertirse en una investigación. Los datos que pueden ser experimentales, de observación, operacionales, de terceros, sector público, seguimiento, procesados e, incluso, reutilizables, los cuales emergen de un proyecto de medición o de una investigación, que genera una cantidad importante de datos para estar vigentes, como se observa en la Figura 3.

LOS PRINCIPIOS RECTORES DE FAIR PARA LA GESTIÓN DE DATOS CIENTÍFICOS

El ecosistema de datos FAIR, según Hodson (2018), se compone por las políticas que regularizan y delimitan los datos, los especialistas que los originan o manejan, los planes de gestión de datos,

Figura 3. Tipo de datos que se trabajan en el RI



Fuente: Desarrollo de un repositorio de datos científicos de apoyo a la investigación: el caso de las ciencias de la tierra (Castro, 2020, p. 88).

los identificadores, los estándares, los repositorios de confianza y los servicios en la red cibernética donde se almacenan los datos. Dentro de todo el ecosistema, este trabajo se centrará, únicamente,

a la parte que refiere a los objetos de datos FAIR, que se compone de su fichero, identificador persistente, descripción de los estándares y formatos utilizados en la obtención, representación, visualización de datos y los metadatos, que describen los *sets* que deben permitir su interpretación y reutilización.

La necesidad urgente de mejorar la infraestructura que respalda a la reutilización de datos académicos y el grupo de personas interesadas, que representan a la academia, la industria, las agencias de financiamiento y los editores académicos, se unieron para diseñar y respaldar los Principios a los que nos referimos, con la intención de utilizarlos como una guía para aquellos interesados que deseen mejorar la reutilización de sus datos. A diferencia de las iniciativas de pares, que se centran en especialistas, estos principios ponen especial énfasis en mejorar la capacidad de las máquinas para encontrar y utilizar los datos automáticamente, además de respaldar su reutilización por parte de los individuos.

Llevar a cabo estos principios logra conectar a los investigadores, editores, bibliotecólogos y repositorios de datos para habilitarlos de la siguiente forma:

- *Recuperable*, los datos pueden ser encontrados por la comunidad después de su publicación, a través de herramientas de búsqueda, mediante la asignación de un identificador único y persistente a los datos y los metadatos.
- *Accesible*, deben poder descargarse por otros investigadores al usar sus identificadores sin necesidad de protocolos especiales.
- *Interoperable*, tanto los datos como los metadatos deben apegarse a estándares abiertos, para permitir su intercambio y reutilización (compatible y, ampliamente, aplicable para representar el conocimiento).
- *Reusable*, los datos y los metadatos pueden ser reutilizados por otros investigadores, al quedar clara su procedencia y las condiciones de reutilización, así como con su licencia sobre el uso y reutilización con la información.

Además de apuntar al descubrimiento científico y mejorar la integridad, transparencia y reproducción de los datos, deberán ser accesibles, interoperables y reutilizables no sólo para las personas, sino también para las máquinas facilitando su descubrimiento, podrán ser citados de manera equivalente a otros resultados académicos mediante protocolos y estándares basados en la web, proporcionando la información de procedencia, licenciamiento de uso e interoperable técnica y semánticamente, así como reutilizables. Los datos también deben estar bien conservados, ser accesibles y estar vinculados de forma segura a las publicaciones asociadas, recursos o facilitar el reuso.

ALCANCE

Es importante señalar que los RI apoyan la diseminación de los recursos de información científica y tecnológica de una entidad y del país, convirtiéndose en un escaparate importante para la producción científica tanto de las publicaciones académicas como de los datos de investigación, para que consigan una mayor visibilidad y desarrollo de nuevas líneas de acción a través de trabajo colaborativo entre especialistas de diferentes disciplinas. Es necesario trabajar junto con otras comunidades de partes interesadas (investigadores, universidades y financiadores de la ciencia) con el propósito de continuar con la misión de ponerlo en abierto y valorar aún más la inclusión, la diversidad, la confianza e innovación.

Como puede observarse, realizar una gestión correcta y adecuada en los repositorios beneficia, en gran medida, la representación de los proyectos de investigación y ofrece un dominio específico o un ámbito donde se requiere comprender el proceso de una exploración a profundidad necesaria para los especialistas. Los investigadores no conocen los esquemas de metadatos ni la lógica del acceso, por lo que nuestra contribución como profesionales de la información es primordial dentro de estas actividades tanto académicas como científicas para seguir en las líneas de la ciencia abierta.

CONSIDERACIONES FINALES

Para concluir es importante mencionar que un repositorio institucional, ofrece un medio de difusión de la ciencia, visibiliza las investigaciones y genera nuevo conocimiento mediante modelos de operación, que cumplan los estándares internacionales con el propósito de enfocarse al almacenamiento, la preservación y la disseminación de la información académica, científica, tecnológica y de innovación, que se derive de las investigaciones en lo que se refiere a datos.

Es imprescindible el fortalecimiento de los Principios FAIR con miras a hacerlos localizables e interoperables mediante metadatos, accesibles y, sobre todo, reutilizables, facilitando el intercambio de la información producida en todas las ramas del conocimiento.

Finalmente, cabe señalar que, al realizar una gestión de la información correctamente en los repositorios se beneficia, en gran medida, la representación de los proyectos de investigación y ofrecen un dominio específico o un ámbito, que se requiere comprender para efectuar un proceso de investigación, que necesitan los especialistas.

BIBLIOGRAFÍA

- Acosta, N. M., Díaz Pérez, M., Giráldez, R. y Quintana, Y. (2018). Política de información para repositorio institucional en universidad cubana: Caso Universidad de Pinar del Río. En J. A. Merlo Vega (Ed.), *Ecosistemas del Acceso Abierto* (pp. 339-351). Ediciones Universidad de Salamanca.
- American Geophysical Union. (2019). Declaration adopted by the American Geophysical Union. <https://ethics.agu.org>.

- Armendáriz, S., Sosa Zaragoza, P. y Sosa Jiménez, M. A. (2021). Los datos para la investigación en los repositorios institucionales y su impacto científico y social: los RI en Ciencias de la Tierra. En M. García, S. M. Del Patrocinio, D. Salas y J. Torres (Coords.), *Experiencias de repositorios institucionales en México* (pp. 153-165). Ediciones Octaedro.
- Castro, M. (2020). “Desarrollo de un Repositorio de datos científicos de apoyo a la investigación: el caso de las ciencias de la Tierra” [Tesis de maestría, Universidad Nacional Autónoma de México]. <http://132.248.9.195/ptd2020/agosto/0802846/Index.html>
- Castro, M., Castro, R. C. y Gómez, C. (2021). Las declaraciones, principios e iniciativas de la Ciencia Abierta: una visión. En M. García, S. M. Del Patrocinio, D. Salas, M. S. Ramírez y J. Torres (Coords.), *Ciencia Abierta opciones y experiencias para México y Latinoamérica* (pp. 79-90). Ediciones Octaedro.
- Conacyt. (2017). *Lineamientos específicos para Repositorios*. <http://bitly.ws/ufAx>
- Conacyt. (2017a). *Lineamientos generales de Ciencia Abierta*. <http://bitly.ws/ufAz>
- Conacyt. (2017b). *Lineamientos jurídicos de Ciencia Abierta*. <http://bitly.ws/ufAE>
- Conacyt. (2019). *El Repositorio Nacional de México*. <http://www.repositorionacionalcti.mx/>.
- Doria, M. V., Prado, A. M. y Haustein, M. C. (2014). *Repositorios digitales de Acceso Abierto: presencia de repositorios en universidades nacionales argentinas*. Editorial Académica Española.
- Drake, M. A. (2004). Institutional Repositories: Hidden Treasures. *Information Today, Inc.*, 12(5), 73-82.

- Ethniko Kai Kapodistriako Panepistimio Athinon. (2009). *Digital Repository Infrastructure Vision for European Research*. <https://cordis.europa.eu/project/id/212147/es>
- Hodson, S. (2018). Making FAIR Data a Reality... and the Challenges of Interoperability and Reusability. En Committee on Data International Science Council, *10th Annual European DDI User Conference, Annual European DDI User Conference*. <https://zenodo.org/record/2530029#.Y9rOonbMLIU>
- IRUSK-UK. (2009). *The Publisher and Institution Repository Usage Statistics* (PIRUS2). <https://irus.jisc.ac.uk/about/pirus2/>
- Lucio Pillasagua, A., Silva Gorozabel, J. J. y Moreira Cañarte, M. K. (2022). “Gestión de la información en las estrategias normativas para la organización y dinámica de la economía de los gobiernos autónomos descentralizados de la provincia de Manabí”. *Serie Científica de la Universidad de las Ciencias Informáticas*, 15(7), 294-305. <https://publicaciones.uci.cu/index.php/serie/article/view/1133>
- LXV Legislatura de la Paridad, la Inclusión y la Diversidad. *Ley de Ciencia y Tecnología*. Nueva Ley publicada en el Diario Oficial de la Federación el 5 de junio de 2002. <http://www.diputados.gob.mx/LeyesBiblio/pdf/LCT.pdf>
- Villanueva Estrada, R., Armendáriz, S., Castro Escamilla, M., Ortiz Guerrero, N. y Rocha Miller, R. (2021). Construcción de un repositorio geotérmico para el desarrollo de recursos geoenergéticos de México. En M. García, S. M Del patrocinio, D. Salas y J. Torres (Coords.), *Experiencias de repositorios institucionales en México* (pp.143-152). Ediciones Octaedro.

Investigación, reflexión, acción desde el Posgrado en Bibliotecología y Estudios de la Información. Instituto de Investigaciones Bibliotecológicas y de la Información-Programa de Posgrado en Bibliotecología y Estudios de la Información/UNAM. Coordinación editorial, Anabel Olivares Chávez; revisión y corrección de pruebas, Diana Serena Palacios; formación editorial, Mario Ocampo Chávez. Primera edición electrónica, 15 de agosto de 2023.