

Explorando el papel de los repositorios de datos de investigación en el contexto de la ciencia abierta

Exploring the role of research data repositories in the context of open science

Eder Ávila Barrientos

UNAM (Instituto de Investigaciones Bibliotecológicas y de la Información)

eder@iibi.unam.mx

eavila86@gmail.com

Resumen

Los repositorios de datos de investigación son una parte importante de la ciencia abierta al propiciar el desarrollo de lineamientos para almacenar, organizar y compartir datos de investigación de manera accesible. Ambos conceptos trabajan juntos para promover una investigación científica más transparente, integral y colaborativa. Se estima que los repositorios de datos de investigación pueden fomentar en mayor grado la reutilización de los datos, esto a través de su organización, recuperación y acceso, generando con ello un entorno interoperable de datos con múltiples significados y tipologías. Mediante este trabajo, se aborda qué son los repositorios de datos de investigación, cuáles son sus propiedades y cuál es su interacción en el contexto de la ciencia abierta, lo cual permita comprender cual ha sido el progreso y desarrollo de la ciencia abierta en función del tsunami de datos que se ha generado a partir de la explosión de las tecnologías digitales y su interacción con la comunidades académicas y científicas.

Palabras clave

Repositorios de datos de investigación; Datos de investigación; Ciencia abierta.

Abstract

Research data repositories are an important part of open science by enabling the development of guidelines for storing, organizing and sharing research data in an accessible way. Both concepts work together to promote more transparent,

comprehensive and collaborative scientific research. It is estimated that research data repositories can promote the reuse of data to a greater extent, through its organization, recovery and access, thereby generating an interoperable data environment with multiple meanings and typologies. Through this work, we address what research data repositories are, what their properties are and what their interaction is in the context of open science, which allows us to understand what the progress and development of open science has been based on the tsunami of data that has been generated from the explosion of digital technologies and their interaction with academic and scientific communities.

Keywords

Research data repositories; Research data; Open Science

Recibido: 29/11/2023

Aceptado: 8/05/2024

DOI: <https://dx.doi.org/10.5557/IIMEI15-N28-001029>

Descripción propuesta: Ávila Barrientos, Eder, 2024. Explorando el papel de los repositorios de datos de investigación en el contexto de la ciencia abierta. *Métodos de Información*, **15**(28), 1-29.

1. Introducción.

La relación entre los repositorios de datos de investigación (RDI) y la ciencia abierta es estrecha y significativa, ya que permite a los investigadores y al público en general acceder a datos de investigación de forma libre o a un costo mínimo. Esto fomenta la transparencia y la accesibilidad de la información científica, lo que es esencial para la ciencia abierta.

La ciencia abierta es un concepto que pone de manifiesto una transformación en la manera de hacer ciencia. De acuerdo con Anglada y Abadal (2018) “la ciencia abierta es esto: un cambio de paradigma en la manera de hacer ciencia. Ésta no cambia sustancialmente con respecto a sus

motivaciones y objetivos, pero sí lo hace (sustancialmente) en cuanto a sus métodos. El cambio no está en lo que se hace, sino en cómo se hace”.

Por su parte la UNESCO (2023) manifiesta que la ciencia abierta es un conjunto de principios y prácticas que tienen como objetivo hacer que la investigación científica de todos los campos sea accesible para todos en beneficio de los científicos y de la sociedad en su conjunto. La ciencia abierta consiste en garantizar no sólo que el conocimiento científico sea accesible sino también que la producción de ese conocimiento en sí sea inclusiva, equitativa y sostenible.

Por lo tanto, la ciencia abierta fundamenta su accionar en el principio de apertura tanto a la información como al conocimiento científico, pero también enfatiza en la manera en cómo esta apertura debe ser asequible por la comunidad científica y la sociedad en general. Al respecto, Beigel (2022) afirma que, aunque el proyecto de ciencia abierta tiene ya algunas décadas de desarrollo, en estos últimos tiempos ha alcanzado una fase de consolidación. Entre sus aceleradores se encuentra la pandemia de la covid-19, que puso sobre el tapete la necesidad de hacer accesible la literatura científica desde los hogares, así como la urgencia de promover formas abiertas de colaboración para que el progreso científico sea un derecho humano y las vacunas estén al alcance de todos.

En este escenario, los datos de investigación que fueron depositados en repositorios de datos de una manera abierta fueron relevantes para desarrollar mecanismos de información y dar seguimiento al fenómeno que provocó la pandemia por COVID-19. Por ejemplo, el *research data produced at Johns Hopkins University* (<https://archive.data.jhu.edu/>) y el proyecto *Coronavirus Resource Center* (<https://coronavirus.jhu.edu/map.html>), fueron fuentes de almacenamiento, organización y visualización de datos en tiempo real, lo que permitió contar con mecanismos validados de información y poder mantener actualizada a la sociedad respecto al avance de la pandemia en diversas latitudes del mundo.

De esta manera, los RDI desempeñan un papel crucial en la preservación a largo plazo de datos científicos, ya que garantizan que los conjuntos de datos estén disponibles de manera permanente, lo que contribuye a la integridad de la investigación y la capacidad de verificar resultados en el futuro.

Además, al poner los datos a disposición del público, los investigadores pueden fomentar la colaboración y la reutilización de datos. Esto puede acelerar el avance del conocimiento y permitir que los datos se utilicen en una variedad de contextos, lo que es fundamental para la ciencia abierta.

Aunado a ello, diversas agencias de financiamiento y revistas científicas requieren que los investigadores depositen sus datos en repositorios públicos como parte de las políticas de ciencia abierta. Se estima que esto garantizaría que los datos estén disponibles para la comunidad científica y académica.

De esta manera, los RDI desempeñan un papel esencial en la promoción de la ciencia abierta al facilitar el acceso abierto a datos de investigación, la preservación de datos y la colaboración entre científicos, lo cual ayuda a garantizar que la investigación científica sea más transparente, reproducible y accesible, lo que, a su vez, contribuye al avance del conocimiento y a una mayor confianza en la investigación científica.

El propósito general de esta investigación es analizar el papel de los repositorios de datos de investigación en el contexto de la ciencia abierta. Para ello, se aborda su interacción con los diferentes actores involucrados en la investigación, así como su contribución a los propósitos de la ciencia abierta. Por su parte, los objetivos específicos consisten en analizar la conceptualización de los repositorios de datos de investigación, identificando sus principales características y funcionalidades; y explorar las propiedades de los repositorios de datos de investigación, en términos de su impacto en la ciencia abierta.

2. Metodología

Se presenta un análisis descriptivo en donde se analiza la interacción de los repositorios de datos de investigación en el contexto de la ciencia abierta. Mediante un proceso de revisión de la literatura, apoyado en el análisis del discurso se examinaron recursos de información vinculados al objeto de estudio planteado en la investigación. Para esto, se analizaron artículos de investigación, capítulos de libro y estudios de caso. Estos recursos de información fueron obtenidos en bases de datos, como Elsevier, Springerlink, IEEE y Emerald; en repositorios institucionales, como el repositorio institucional de la UNAM; además, en sitios web gubernamentales, académicos e institucionales especializados en el objeto de estudio planteado en la investigación (Research Data Alliance, Research Data Management @Harvard).

La selección de estas fuentes responde a su especificidad temática relacionada con los repositorios de datos de investigación, pues son fuentes que tratan de manera particular el estudio de los repositorios desde diversas aristas, por ejemplo, procedimental, teórica, pragmática y tecnológica. Dentro de estas fuentes, se implementaron estrategias de búsqueda relacionadas con los siguientes términos: repositorios datos de investigación, ciencia abierta, datos de investigación y sus equivalentes en inglés.

Cabe señalar que, aunque los términos empleados permitieron recuperar recursos que utilizaran dichos términos dentro de su título, fue necesario revisar detalladamente cada recurso para identificar el enfoque disciplinario al que estaban dirigidos. Pues la noción conceptual de datos representa a múltiples variables de estudio, desde aquellas que versan en el rubro técnico y pragmático, hasta aquellas que permiten considerar sus nociones epistemológicas. En el caso de los sitios de Research Data Alliance y Research Data Management @Harvard, no se realizaron búsquedas específicas de recursos, se llevó a cabo

un análisis de la documentación que contenían en sus páginas web que trataran los temas planteados en el objeto de estudio de la investigación.

Por lo tanto, se seleccionaron recursos publicados de 2016 a 2023, tomando en cuenta criterios que reflejan una condicionante tecnológica y cronológica, pues el estudio de los datos se trata de un fenómeno que se encuentra en constante dinamismo debido a la aparición de múltiples publicaciones que analizan su impacto en diferentes contextos y disciplinas. Además, que la mayoría de los recursos utilizados fueron publicados en idioma inglés.

La periodicidad de los años seleccionados responde a la constante actualización de los contenidos y tópicos que se presenta en el objeto de estudio de los datos, pues en años recientes la literatura que versa sobre este tema ha crecido exponencialmente. Los recursos de información publicados fuera del rango periódico mencionado, fueron elegidos tomando en cuenta su grado de contribución y pertinencia respecto al análisis del objeto de estudio planteado en este trabajo. Es decir, tomando en cuenta su relevancia para obtener una visión global del tema y que se relacionara con los tópicos de la sistematización y organización de los datos de investigación.

3. ¿Qué son los repositorios de datos de investigación?

Los repositorios de datos de investigación son sistemas que permiten el almacenamiento, organización y acceso a conjuntos de datos científicos y académicos. Estos repositorios se utilizan para preservar y compartir datos de investigación, lo cual resulta fundamental para la ciencia abierta y la transparencia de la investigación.

Desde hace más de una década los RDI han tenido una notable presencia en el ámbito académico y científico. Estos sistemas para organizar, buscar,

recuperar y acceder a conjuntos de datos han sido desarrollados mediante principios y directrices internacionales como, OpenAire (<https://www.openaire.eu/about>) y FAIR (<https://www.go-fair.org/fair-principles/>), que tienen el propósito de fomentar la interoperabilidad en esquemas, modelos y los propios datos que son organizados y sistematizados.

De acuerdo con Ma (2022), el contenido almacenado y servido por repositorios de datos incluye conjuntos de datos digitales y catálogos de datos.

Por su parte, Pampel, et al. (2013) afirman que los RDI son sistemas que almacenan una amplia variedad de formatos de archivo bajo diferentes condiciones de acceso y reutilización, las cuales pueden ser abiertas, restringidas y cerradas. Cabe señalar que este tipo de repositorios respaldan el acceso a los datos mediante el uso de licencias, por ejemplo, la licencia Creative Commons CC BY 4.0 (CC Atribución) y la licencia CC0 (Creative Commons Zero).

Además, los identificadores persistentes como DOI, ORCID, ISNI, VIAF y DAI permiten la localización e interoperabilidad de los datos en un contexto digital común. Por su parte, los esquemas de metadatos como Dublin Core, DataCite y Data Documentation Initiative, hacen posible la descripción, representación y preservación a largo plazo de los datos de investigación. Bajo esta premisa, las interfaces usables que acompañan a dichos repositorios fomentan el acceso entre varios depósitos de datos de una manera integral y estructurada.

En general, los RDI se pueden clasificar en diferentes tipos según los datos y contenidos informativos que organizan. Entre esos tipos, destacan aquellos que ofrecen acceso a datos digitales y un servicio de catálogo de datos que permite consultar los metadatos que describen y gestionan a los datos, dentro de una determinada institución o contexto.

De esta manera, los RDI son de diversa naturaleza y tipología. Destacan aquellos que presentan las siguientes características:

- RDI de carácter institucional, los cuales almacenan los datos generados por comunidades de académicas y científicas de una determinada universidad o institución académica dedicada a la investigación. Un ejemplo de este tipo de repositorio es el de la Universidad de Múnich (<https://data.ub.uni-muenchen.de/>).
- De carácter disciplinario, en donde se almacenan conjuntos de datos de una temática o asunto en específico relacionado con el desarrollo de investigación básica o aplicada, como por ejemplo, el repositorio GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) desarrollado por el National Center for Biotechnology Information de la Biblioteca Nacional de los Estados Unidos.
- También pueden localizarse repositorios de carácter multidisciplinario, los cuales contienen conjuntos de datos relacionados con el desarrollo de investigación en diversas áreas del conocimiento o disciplinas. El repositorio de datos Figshare (<https://figshare.com/search?q=open%20research%20data%20>) es un ejemplo de este tipo de repositorios.
- Por otra parte, se puede acceder a repositorios relacionados con el desarrollo de proyectos específicos de investigación, por ejemplo, el repositorio del Bern Digital Pantheon Project (<http://repository.edition-topoi.org/collection/BDPP>). En este repositorio, además de acceder a repositorios de datos, el usuario puede interactuar con una serie de visualizaciones y descripciones de imágenes relativas al análisis de datos llevado a cabo como parte de dicho proyecto, tal como se aprecia en la figura 1. Aunado a ello, puede observarse el papel que la visualización de datos tiene en el acceso y recuperación que se ejerce en este tipo de repositorios, pues se requieren de mecanismos interactivos e intuitivos para que el usuario pueda interactuar con los

datos y conocer sus propiedades y navegar por contenidos audiovisuales relacionados con los propios datos que se representan en el repositorio.

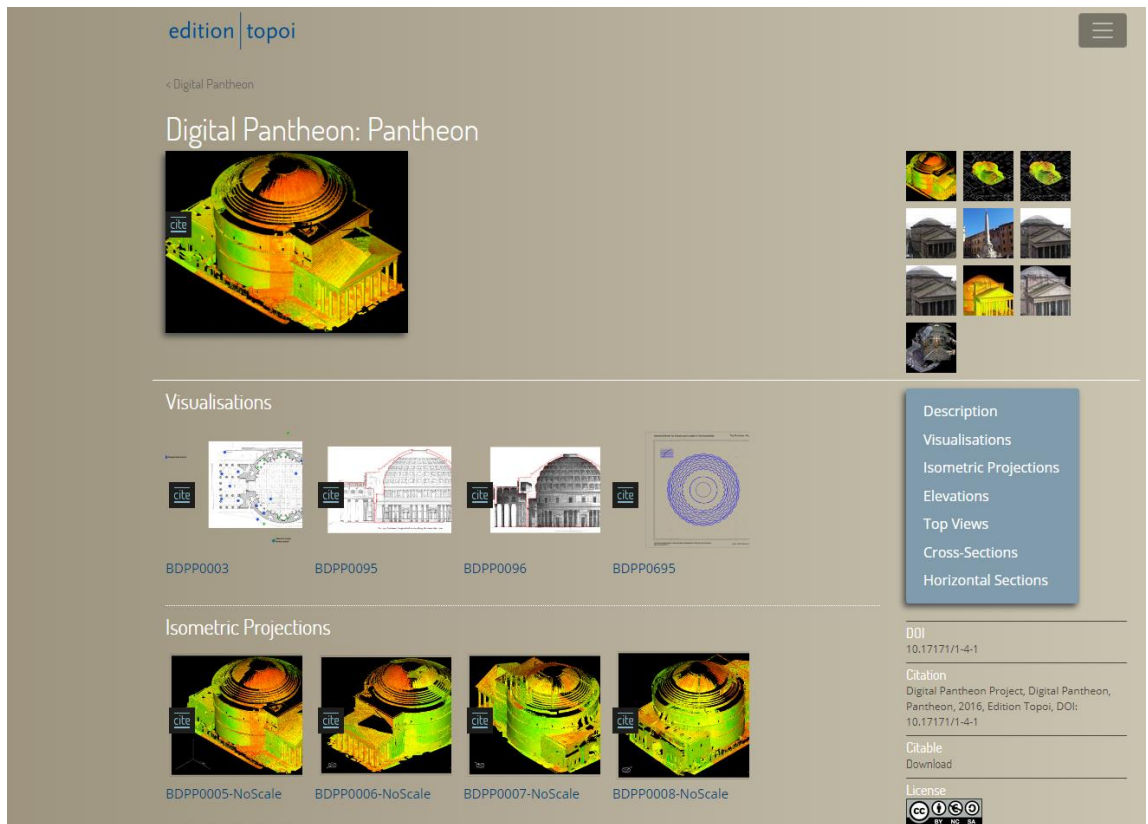


Figura 1. Digital Pantheon Project. 2023. Fuente:
<http://repository.edition-topoi.org/collection/BDPP/object/1>

Por otra parte, los planes para la gestión de datos de investigación se guían con base al concepto de trazabilidad de los datos, en donde se busca transparentar el uso y manejo de los datos desde su creación, hasta su alcance en la publicación de resultados que forman parte de las investigaciones iniciales y que tienen una relación directa con los principios institucionales para el desarrollo de dicha investigación. En este escenario es en donde se puede apreciar la presencia de las hojas de ruta, las cuales permiten trazar el cumplimiento de los objetivos marcados en un determinado proyecto de investigación.

Según CODATA (2022), la trazabilidad es un concepto que interactúa con el ciclo de vida de los datos para identificar todos los accesos y cambios realizados a los datos en un determinado sistema o contexto, con lo cual ayuda a demostrar transparencia, cumplimiento y adherencia a las directrices que permean a los datos de investigación.

De esta manera, la trazabilidad de los datos, junto con la documentación de los datos, puede considerarse parte de un proceso de auditoría, en donde la trazabilidad de los datos es fundamental para una investigación reproducible y sobre todo reutilizable. De esta manera, dichos planes contemplan que los datos de investigación puedan estar accesibles y ser difundidos mediante un sistema desarrollado expreso para este propósito, como pueden ser los RDI.

La normatividad relativa y aplicable a los datos de investigación en el contexto español, puede consultarse en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo (RGPD) publicado el 27 de abril de 2016 en el Diario Oficial de la Unión Europea. En este documento se expresa la importancia de anonimizar y seudonimizar datos de investigación que tengan carácter personal. Pues los datos personales pueden ser sensibles y expuestos a malos usos. En este sentido, la Unión Europea recomienda que los archivos de datos personales destinados a investigación sean principalmente archivos anonimizados, con un doble fin: evitar la compleja gestión en materia de protección de datos y favorecer la libre circulación de datos anonimizados en el ámbito de los datos masivos (Parlamento Europeo y Consejo de la Unión Europea, art. 26).

A su vez, la seudonimización de los datos personales es una práctica muy habitual en el contexto de las Ciencias de la Salud y la Biomedicina, por lo que los conjuntos de datos seudonimizados, y la información adicional vinculada con dichos conjuntos, estarán bajo el ámbito de aplicación del RGPD, así como el tratamiento que los genera. Ofreciendo con ello un marco común

normalizado para el uso y tratamiento de datos personales en el ámbito de la investigación.

Si bien, la normatividad aplicable a los datos de investigación, no ofrece un panorama general de aplicación, si es posible apreciar diversas normativas e instrumentos que tiene injerencia sobre todo en aspectos específicos como puede ser la protección de datos personales. Esto responde a la naturaleza y dinamismo que los propios datos presentan, aspectos que resaltan al momento de organizarlos y sistematizarlos a través de un repositorio.

4. Propiedades de los repositorios de datos de investigación

Los RDI tienen características que forman parte de su estructura y permiten organizar, buscar, recuperar y acceder a datos de investigación. Al respecto, los datos abiertos de investigación se refieren a la práctica de hacer que los datos generados en el curso de investigaciones científicas y académicas estén disponibles públicamente para su acceso y uso por otros investigadores, profesionales, estudiantes y el público en general.

Se estima que estas acciones pueden fomentar la transparencia, la colaboración y la reutilización de estos datos en diversos ámbitos disciplinarios. Por ejemplo, los RDI están siendo desarrollados por instituciones y organizaciones alrededor del mundo, para ello han establecido repositorios donde los investigadores pueden cargar y compartir sus datos de una manera organizada y sistematizada.



Figura 2. Aspectos de un repositorio de datos de investigación con los íconos correspondientes utilizados en re3data.org. Fuente:

<https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0078080.g001>

Las propiedades de los RDI pueden resumirse en una serie de criterios que tienen que ver con su estructura y elementos que permiten su funcionalidad. Por ejemplo, en la figura 2 pueden observarse los aspectos que un RDI tiene al momento de concebir su arquitectura de una manera integral.

Al respecto, el proyecto R3Data (<https://www.re3data.org/>) es un registro global de repositorios de datos de investigación que integra repositorios de diferentes disciplinas académicas. Presenta una plataforma para el almacenamiento permanente y el acceso a conjuntos de datos para investigadores, organismos de financiación, editores e instituciones académicas y tiene como objetivo promover una cultura de intercambio, mayor acceso y mejor visibilidad de los datos de investigación (German Research Foundation, 2023). R3Data utiliza una gama de iconos para identificar si un RDI que aspire a ser integrado y listado en su directorio, cumple cabalmente con los criterios que se solicita.

En términos generales, un RDI está respaldado por el uso de estándares técnicos como los identificadores persistentes; la presencia de aspectos legales que son cubiertos mediante el uso de licencias; propiciar diferentes tipos de acceso a los datos, los cuales pueden ser abierto, restringido o cerrado; desarrollar una política de uso que respalde la funcionalidad del repositorio; incluir documentación en información general acerca del uso y organización del repositorio de datos; y ofrecer a través de estándares de calidad, certificaciones relacionadas con la descripción y representación de los datos.

En este sentido, los metadatos y los RDI se relacionan en el contexto de la gestión de datos, especialmente en el ámbito de la investigación y la ciencia, pues los metadatos desempeñan un papel crucial en la organización efectiva de los datos en este tipo de repositorios.

Repository details

Harvard Dataverse

General Institutions Terms Standards

Name of repository	Harvard Dataverse
Additional name(s)	The Dataverse Project
Repository URL	https://dataverse.harvard.edu/
Subject(s)	Social Sciences Economics Astrophysics and Astronomy Basic Biological and Medical Research Social and Behavioural Sciences Humanities and Social Sciences Physics Natural Sciences Biology Life Sciences
Description	The Harvard Dataverse is open to all scientific data from all disciplines worldwide. It includes the world's largest collection of social science research data. It is hosting data for projects, archives, researchers, journals, organizations, and institutions.
Content type(s)	Standard office documents Databases Scientific and statistical data formats Raw data Archived data Source code Software applications
Keyword(s)	FAIR software research demography epidemiology human societies human behavior multidisciplinary social societies
Persistent identifier(s) of the repository	RRID:SCR_001997 RRID:nif-0000-00316 FAIRSharing_doi:10.25504/fairsharing.12e1ss
Repository size	4,749 dataverses, 108,231 datasets, 953,876 files
Repository type(s)	disciplinary institutional
Mission statement for designated community	https://dataverse.org/about
Research data repository language(s)	English
Data and/or service provider	data provider service provider

Back to search Submit a change request Get a badge

Cite this re3data.org record:
re3data.org: Harvard Dataverse; editing status 2021-09-03; re3data.org - Registry of Research Data Repositories.
<http://doi.org/10.17616/R3C880> last accessed: 2023-11-24

Figura 3. Ejemplo de un registro de metadatos para la descripción de un repositorio de datos de investigación. Fuente:

<https://www.re3data.org/repository/r3d100010051>

Como puede apreciarse en la figura 3, los metadatos se utilizan para describir a los conjuntos de datos almacenados en un repositorio. Cada conjunto de datos en un repositorio suele estar acompañado de metadatos que proporcionan información esencial sobre el contenido y tipología de los datos. Esto permite a los usuarios entender de qué se trata un conjunto de datos sin necesidad de acceder a los datos subyacentes de inmediato.

De acuerdo con Farnel y Shiri (2014), los metadatos brindan información estructurada que proporciona contexto para objetos de información de todo tipo, incluidos los datos de investigación y, al hacerlo, permite el uso, la preservación y la reutilización de esos objetos.

Además, los metadatos facilitan la búsqueda, recuperación y el descubrimiento de datos en los repositorios. De esta manera, los usuarios pueden utilizar los metadatos para buscar conjuntos de datos relevantes o filtrar los resultados de búsqueda según criterios específicos, como el autor, la fecha o las palabras clave.

Por lo tanto, los metadatos también desempeñan un papel crucial en la interoperabilidad entre diferentes repositorios de datos. Al seguir estándares de metadatos comunes, los datos se pueden compartir y reutilizar de manera más efectiva entre diferentes sistemas y organizaciones. Los metadatos pueden incluir información sobre el acceso y la utilización de los datos, lo que es importante para el cumplimiento de regulaciones y políticas éticas, como el acceso abierto, la privacidad y la atribución adecuada.

Los metadatos son elementos que proporcionan un contexto acerca del alcance de los conjuntos de datos que son depositados y descritos en un repositorio, ya que proporcionan detalles que facilitan el acceso y la comprensión de los datos de investigación, lo cual es fundamental en el entorno de la investigación científica y académica. Pues en la actualidad, se requiere de sistemas que propicien una búsqueda y recuperación de datos mucho más

intuitiva e interactiva, acorde al dominio temático y a la naturaleza de los datos que serán utilizados.

5. Desarrollo de repositorios de datos de investigación

El desarrollo de estos sistemas puede constatarse en plataformas de amplio alcance que dentro de sus estructuras permiten acceder a datos de investigación con diferentes características y propósitos. Por ejemplo:

- Dryad (<https://datadryad.org/stash>). Es una plataforma de publicación de datos abiertos, que representa a una comunidad comprometida con la disponibilidad abierta y la reutilización de los datos de investigación en un alcance internacional. En la actualidad almacena conjuntos de datos multidisciplinarios de grupos de investigación de diversas instituciones académicas alrededor del mundo. Entre julio de 2020 y junio de 2021, Dryad había publicado cerca de 7,381 conjuntos de datos en su plataforma (Dryad, 2022).

- Zenodo (<https://zenodo.org/>). Es un proyecto dirigido desde el centro de datos del CERN, cuyo fin es la preservación a largo plazo de datos de investigación y objetos digitales. El CERN mantiene uno de los conjuntos de datos científicos más grandes del mundo sobre física de altas energías a través de Zenodo. En la actualidad esta plataforma recibe financiación por parte de la Comisión Europea a través de la iniciativa Open Aire.

- Mendeley Data (<https://data.mendeley.com/>). Es una plataforma para el acceso a datos de investigación, la cual se encuentra fundamentada en los principios de *Digital Commons Data*, lo cual permite proporcionar acceso abierto o restringido a los datos con identificadores únicos persistentes para permitir su referenciación y citación. Este proyecto se integra con portales de revistas científicas como

ScienceDirect, cellPres, lo cual permite vincular a los datos con sus respectivos creadores y productos de investigación realizados.

- CORA (<https://dataverse.csuc.cat/>). Es un repositorio de datos de investigación federado y multidisciplinar que permite a las universidades catalanas, a los centros de investigación de Cataluña y a otras entidades de investigación, publicar conjuntos de datos de investigación siguiendo los principios de FAIR y las directrices de la European Open Science Cloud (EOSC). En la actualidad este repositorio cuenta con 765 conjuntos de datos de investigación, entre los que destacan aquellos conjuntos relativos a las ciencias sociales (153 conjuntos de datos = 19% del total del repositorio), artes y humanidades (141 conjuntos de datos = 17% del total del repositorio), medicina y ciencias de la salud (125 conjuntos de datos = 15% del total del repositorio). Cabe señalar que los tipos de datos que se concentran en este repositorio pueden ser aquellos relacionados con el desarrollo de aplicaciones informáticas, audio, imagen, modelos, texto y video. A la fecha, este repositorio tiene un total de 73,849 descargas de conjuntos de datos (CORA, 2023).

- Eciencia.datos (<https://edatos.consorciomadrono.es/>). Se trata de un repositorio de datos de investigación que se creó en diciembre de 2016 con el objetivo de dar visibilidad a los datos de investigación de las universidades miembro del Consorcio Madroño, garantizar su preservación y facilitar su acceso y reutilización (Universidad Autónoma de Madrid, 2023). Actualmente este repositorio cuenta con 1018 conjuntos de datos, de los cuales destacan aquellos relativos al área de las ciencias sociales con 419 conjuntos de datos (mismos que representan el 38% del total del repositorio); 416 (38% del repositorio) referentes a las artes y humanidades y 71 conjuntos de datos (6.5% del total del

repositorio) que están relacionados con el ámbito de las ciencias de la tierra y ciencias ambientales.

En estos desarrollos pueden observarse un proceso de intercambio de datos motivado por el principio de apertura a la información y el seguimiento de principios de la ciencia abierta. Por lo que la presencia de los datos abiertos de investigación juega un papel fundamental en el desarrollo de repositorios de datos de investigación los cuales son organizados mediante una estructura basada en principios de accesibilidad, usabilidad e interoperabilidad.

6. Interacción de los RDI en el contexto de la Ciencia Abierta

La ciencia abierta es un enfoque de investigación que promueve la transparencia, la colaboración y el acceso libre a los resultados y productos de la investigación científica. Fue concebida para mejorar la calidad y la accesibilidad de la ciencia, y fomentar una mayor participación de la comunidad científica y el público en general.

En el contexto español, la legislación aplicable en ciencia abierta, pueden observarse en la Estrategia Nacional de Ciencia Abierta 2023-2027, la cual recoge todos los compromisos relativos a la ciencia abierta adoptados por distintos agentes públicos del sistema, los incluidos en la reforma de la Ley de la Ciencia, la Tecnología y la Innovación, aprobada en septiembre de 2022, Ley 17/2022; la Estrategia Española de Ciencia, Tecnología e Innovación 2021-2023; y el Plan Estatal de Investigación Científica, Tecnológica e Innovación 2021-2022 (Ministerio de Ciencia, Innovación y Universidades, 2023).

Por otra parte, la Nube Europea de Ciencia Abierta (EOSC) apoya a la gestión y aplicación de datos de investigación para garantizar el acceso de los científicos a la ciencia basada en datos. “La EOSC se fundamenta en la Política de Ciencia Abierta de la Unión Europea y en la Estrategia Europea de Datos”

(EOSC, 2023). Con lo cual puede observarse un marco normativo relativo a la Ciencia Abierta que respalda las acciones encaminadas a la apertura de datos de investigación en el contexto europeo.

De acuerdo con Vicente y Martínez (2018) la ciencia abierta es un fenómeno disruptivo que está surgiendo en torno al mundo y especialmente en Europa. La ciencia abierta genera cambios socioculturales y tecnológicos, basados en la apertura y la conectividad, en cómo se diseña, realiza, captura y evalúa la investigación.

Según Abadal (2021), esta apertura de los datos de investigación tiene como objetivo fundamental facilitar su reutilización por parte de otro personal investigador; va así un poco más allá que el acceso abierto centrado especialmente en la difusión de los contenidos.

Aunado a ello, la ciencia abierta promueve la reproducibilidad de los resultados científicos, lo que significa que otros investigadores tendrían la capacidad de repetir un estudio utilizando los mismos datos y métodos. La apertura de datos facilita este proceso al permitir que otros examinen los datos originales y confirmen o refuten los hallazgos de un estudio.

De esta manera, se estima que la ciencia abierta puede fomentar la colaboración entre investigadores de diferentes instituciones y áreas de estudio, pues la apertura de datos permite compartir datos entre colaboradores de manera eficiente, lo que puede acelerar la investigación y llegar descubrimientos más sólidos.

Por otra parte, la ciencia abierta también se preocupa por la accesibilidad de los resultados de la investigación para el público en general. La apertura de datos contribuye a que las personas ajenas a la comunidad científica puedan acceder a información relevante y comprender mejor los avances científicos. Así pues, la ciencia abierta promueve la transparencia en todo el proceso de investigación, desde la recopilación de datos hasta la publicación de los mismos.

Entonces, la apertura de datos es una forma de lograr esta transparencia al permitir que otros pares examinen los datos subyacentes y evalúen la integridad de un determinado estudio.

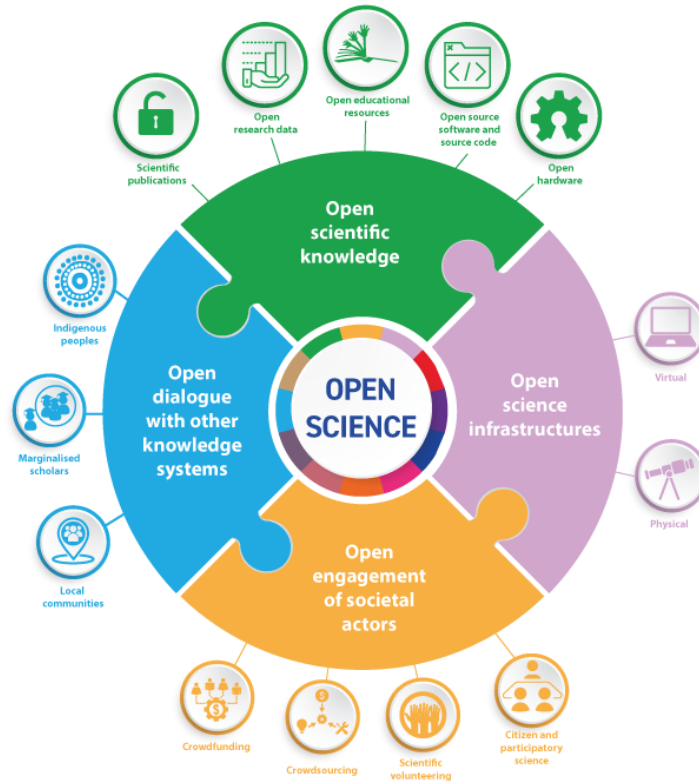


Figura 4. Entendiendo el movimiento de ciencia abierta. Fuente: UNESCO.org, 2023. CC BY-SA 4.0,

<https://commons.wikimedia.org/w/index.php?curid=130456946>

Como puede apreciarse en la figura 4, la ciencia abierta y la apertura de datos son elementos que se encuentran estrechamente relacionados, ya que ambos buscan mejorar la calidad y la confiabilidad de la investigación científica al promover la transparencia, la colaboración y el acceso abierto a los datos de investigación. Estos principios son fundamentales para el avance de la ciencia y para asegurar que los alcances de la investigación sean ampliamente compartidos. No obstante, el tema de la ciencia abierta ha puesto sobre la mesa

cuestiones relativas a la reutilización de los datos para verificar los resultados de la investigación, lo cual fomentaría la aceleración del conocimiento científico, dando pauta con ello a la generación de un cuarto paradigma fundamentado en el descubrimiento intensivo basado en datos. Al respecto, Parastatidis (2009) afirma que, mediante el uso de la tecnología y la automatización de procesos, se ha intentado hacer frente a los desafíos de la avalancha de datos que acontece actualmente en el ambiente digital, por lo que el surgimiento de la Web como plataforma de aplicación, intercambio de datos y colaboración ha derribado algunas barreras para propiciar una mayor difusión y visibilidad de la investigación. Bajo esta premisa, la ciencia abierta se relaciona estrechamente con la apertura de datos en varios aspectos, por ejemplo, implica hacer que los datos de investigación estén disponibles públicamente para que otros investigadores puedan acceder, reutilizar y verificar los resultados de un estudio. La idea es que los datos subyacentes a una investigación deben ser tan accesibles como sea posible, permitiendo que otros científicos construyan sobre ellos y, en última instancia, mejoren la calidad de la ciencia.

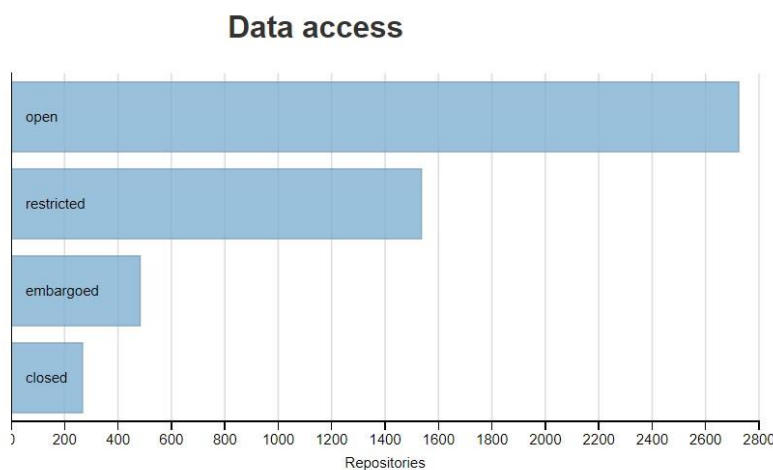


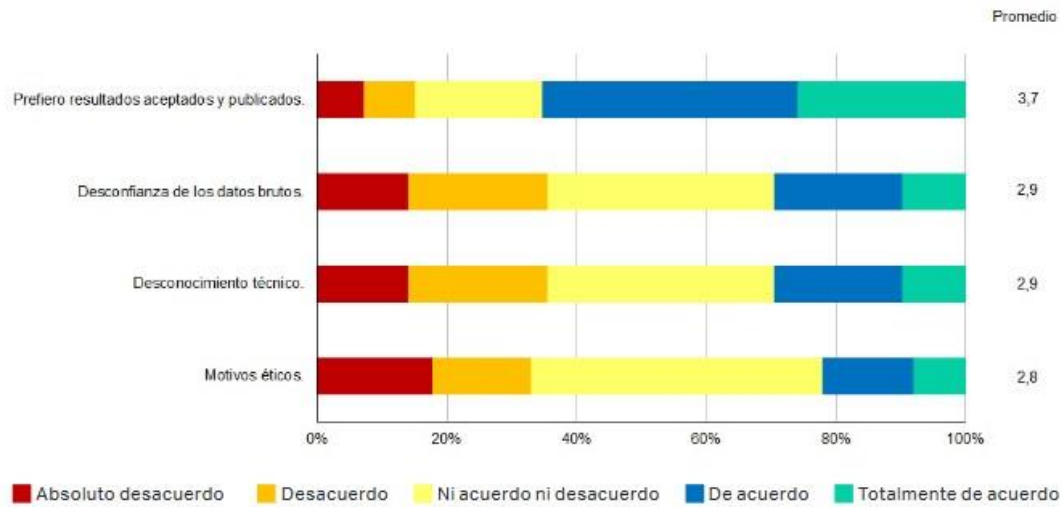
Figura 5. Acceso a los datos. Fuente: R3data Metrics. 2023. Disponible en: <https://www.re3data.org/metrics/dataAccess>

Como puede apreciarse en la figura 5, de acuerdo con las métricas realizadas por la plataforma R3Data, la apertura de los datos es una condición que muestra tendencia significativa en la actualidad, pues de los 2,800 repositorios de datos de investigación que tienen registrados, 2, 724 son de acceso abierto; 1536 de acceso restringido; 483 se encuentran en periodo de embargo y solo 267 son cerrados. Estas cifras permiten obtener una visión global del estado de los repositorios de datos en un escenario integral, que está caracterizado por la generación de políticas de acceso abierto a los datos en diversas instituciones académicas y científicas en el mundo, además de conocer una reconfiguración de los procesos de comunicación científica, que muestra una inclinación por la apertura de los recursos de información a nivel internacional.

No obstante, Borgerud y Borglund (2020) afirman que existen argumentos a favor y en contra de los datos abiertos de investigación por parte de la comunidad científica y académica. Por una parte, la rentabilidad y la democratización son dos aspectos que permiten a estos datos estar accesibles abiertamente para los académicos e investigadores, dando oportunidad de reutilizarlos, lo que aumenta la transparencia y visibilidad de la investigación. Por el contrario, se aprecia una falta de conocimiento sobre la gestión de datos abiertos, tanto por parte de los investigadores como de las instituciones, lo cual tiene un impacto en la manera de concebir la apertura y la reutilización de los datos de investigación por parte de las comunidades científicas y académicas.

De acuerdo con Abadal et al (2023, 39), los principales motivos para no compartir los datos de investigación tienen que ver con la no obligatoriedad para hacerlo, con aspectos relacionados con la privacidad y la confidencialidad, y con el desconocimiento sobre la compartición de datos. En este sentido, la reutilización de los datos de investigación es un proceso que puede ser confiable solo si se utilizan fuentes validadas de información, como es el caso de las publicaciones académicas y científicas. Como puede observarse en la figura 6,

la opinión de los investigadores al momento de reutilizar datos se encuentra fundamentada en la revisión por pares que deben tener estos datos al momento



de publicarse.

Figura 6. Motivos para no usar los datos, según la respuesta de investigadores. Fuente: Ciencia abierta en España 2023: informe de situación y análisis de la percepción, disponible en:

https://diposit.ub.edu/dspace/bitstream/2445/200020/5/CA_informe%207.7.23b.pdf

Por otra parte, el informe The State of Open Data (2023, 21) ha revelado que las motivaciones para compartir datos de investigación son en gran medida consistentes entre los académicos que publicaron su primer artículo entre 1980 y 2023.

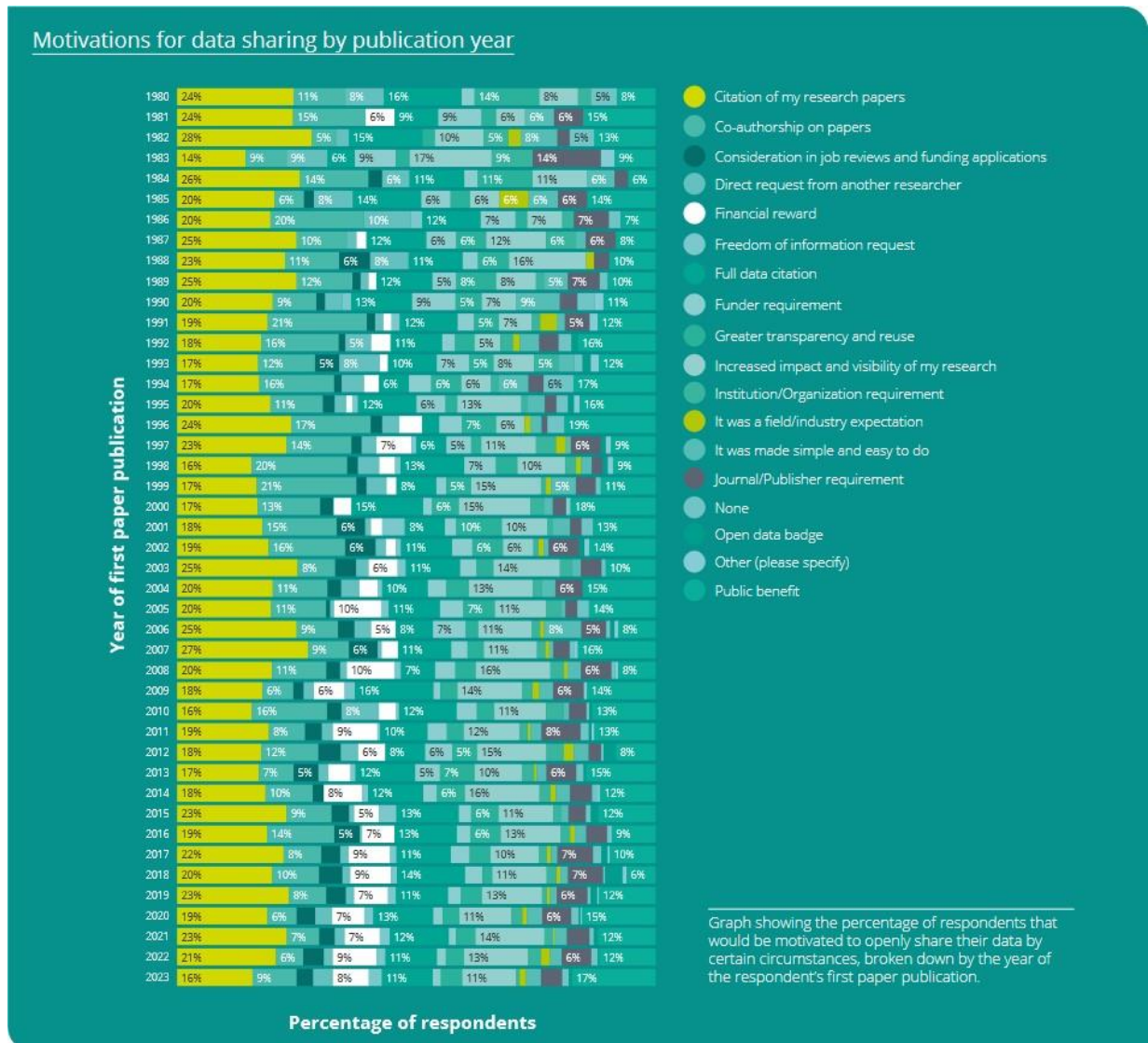


Figura 7. Motivaciones para compartir datos por año de publicación.

Fuente: The State of Open Data 2023. Disponible en:

[https://digitalscience.figshare.com/articles/report/The State of Open Data 2023/24428194](https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2023/24428194)

Como se puede apreciar en la figura 7, cada línea del gráfico representa uno de los grupos tomados en cuenta para el desarrollo del informe. Si bien existe cierta variación año tras año, no se observa ninguna tendencia general. Un mensaje prevalece en todas las etapas de la progresión de la carrera académica: "Los investigadores reciben crédito insuficiente por compartir sus datos abiertamente" (Hahnel et al., 2023).

Junto con las motivaciones para compartir la investigación, parece que la comunicación con los investigadores al mismo nivel sobre cómo obtener más crédito por su investigación y cómo compartir abiertamente los datos de su investigación puede ayudar a mejorar la confianza en torno a sus trabajos de investigación y generar más citas de su artículo.

Debido a lo anterior, se debe tomar en consideración que al utilizar datos de investigación es importante comprender y respetar cualquier restricción de uso o licencia asociada con los datos. Además de citar adecuadamente los datos que se utilizan en determinadas investigaciones, dando crédito a los investigadores originales. Como parte de estas acciones se han generado propuestas para facilitar la citación y la descripción de datos de investigación, como es el caso de la iniciativa DataCite Metadata Schema (<https://schema.datacite.org/>).

De esta manera, la disponibilidad y el acceso a datos abiertos de investigación puede variar según la disciplina y la ubicación geográfica en donde se contemplen. Por ello es preciso tomar en consideración las políticas, procesos y modelos de comunicación científica que tienen una interacción directa con la manera de hacer investigaciones en las diversas instituciones académicas y científicas a nivel global, pues la apertura de los datos y la ciencia abierta no es una constante en todas las latitudes del mundo.

En este contexto, los repositorios de datos de investigación forman parte del engranaje de la ciencia abierta a través de atender la necesidad de describir,

organizar, recuperar y acceder a dichos datos de una manera sistematizada. Pues la primera condicionante para reutilizar a los datos de investigación de una manera libre y abierta, consiste en dotar de una mayor apertura y accesibilidad a los mismos.

7. Consideraciones finales

Los repositorios de datos de investigación son sistemas que permiten almacenar, organizar, recuperar y acceder a conjuntos de datos de diversa naturaleza temática y tipología. En el ámbito de la investigación científica y académica, estos repositorios fomentan la reutilización de los datos y motivan el desarrollo de la ciencia abierta.

Los datos de investigación son almacenados en los RDI para propiciar su reutilización y accesibilidad. Sin embargo, no todas las comunidades académicas y científicas son parte de esta tendencia, pues a nivel internacional se muestra un desarrollo paulatino de la implementación de repositorios de datos de investigación, tanto en universidades como en institutos o centros de investigación.

La implementación de repositorios de datos de investigación a nivel internacional presenta desafíos y consideraciones adicionales debido a la naturaleza global de los datos y las implicaciones legales, culturales y tecnológicas. Aunado a ello, el desarrollo de repositorios de datos de investigación a nivel internacional implica desafíos particulares relacionados con la colaboración, el intercambio de datos y el cumplimiento normativo.

Para facilitar el descubrimiento y la interoperabilidad de los datos de investigación, se emplean estándares de metadatos reconocidos a nivel internacional, como Dublin Core, DataCite o Schema.org. Esto asegura que

otros investigadores puedan comprender y utilizar los datos de manera efectiva, a través de su búsqueda, recuperación y accesibilidad.

A su vez, las políticas de licencias para el uso de los datos de investigación permiten definir cómo los investigadores pueden utilizar y compartir los datos, tomando en cuenta sus condiciones de apertura y reutilización, dos elementos importantes para el desarrollo de la ciencia abierta.

De esta manera, mediante el desarrollo de políticas para la apertura de los datos, se debe promover y respetar las leyes de derechos de autor y propiedad intelectual en diferentes países, pues este marco asegura que los datos compartidos cumplan con las restricciones legales aplicables.

Dado que los datos de investigación pueden estar sujetos a regulaciones de privacidad y protección, se debe tomar en cuenta las leyes de protección de datos de todas las jurisdicciones correspondientes. Esto es especialmente importante cuando se comparten datos de investigación que contienen información personal.

Por lo tanto, si los datos de investigación involucran a personas, se debe garantizar la privacidad y el consentimiento informado de acuerdo con las regulaciones éticas y de privacidad en la investigación. Finalmente, la implementación de repositorios de datos debe considerar la aplicabilidad de medidas sólidas de seguridad y almacenamiento de datos para proteger la integridad de los datos de investigación, especialmente si contienen información sensible.

8. Agradecimientos

Artículo realizado gracias al Programa UNAM-DGAPA-PASPA Convocatoria 2023.

Al Dr. Andreu Sulé Duesa profesor-investigador de la Facultad de Información y Medios Audiovisuales de la Universidad de Barcelona por su orientación que fue fundamental para el desarrollo de este proyecto.

9. Obras consultadas

ABADAL, E., 2021. Ciencia abierta: un modelo con piezas por encajar. *Arbor*, 197(799), 1-12. Disponible: [10.3989/arbor.2021.799003](https://doi.org/10.3989/arbor.2021.799003).

ABADAL, E., et al., 2023. *Ciencia abierta en España 2023: informe de situación y análisis de la percepción*. Barcelona-València: Grupo Ciencia Abierta. <http://hdl.handle.net/2445/200020>

ANGLADA, L. y ABADAL, E., 2018. ¿Qué es la ciencia abierta? *Anuario ThinkEPI*, 12, 292-298. Disponible en: <http://10.3145/thinkepi.2018.43>

BEIGEL, M., 2022. El proyecto de ciencia abierta en un mundo desigual. *Relaciones Internacionales*, 50, 163-181. Disponible: <http://10.15366/relacionesinternacionales2022.50.008>

BORGERUD, C.; BORGLUND, E., 2020. ¿Open research data, an archival challenge? *Archival Science*, 20(3), 279-302. [10.1007/s10502-020-09330-3](https://doi.org/10.1007/s10502-020-09330-3).

CODATA, 2023. Data traceability. *The Committee on Data for Science and Technology* [consulta noviembre 3, 2023]. Disponible en: <https://codata.org/rdm-terminology/data-traceability/>

CORA, 2023. *Repositori de dades de recerca*, [consulta noviembre 18, 2023]. Disponible en: <https://dataverse.csuc.cat/statistics.xhtml>

DRYAD, 2022. *Governance annual reports at main*, [consulta noviembre 9, 2023].

Disponible en:

<https://github.com/datadryad/governance/tree/main/annual-reports>

FARNEL, Sharon y SHIRI; Ali, 2014. Metadata for Research Data: Current Practices and Trends. *International Conference on Dublin Core and Metadata Applications*. pp. 74-82.

GERMAN RESEARCH FOUNDATION, 2023. *Registry of Research Data Repositories*. DOI [10.17616/R3D](https://doi.org/10.17616/R3D).

HAHNEL, M.; SMITH, G.; SCAPLEHORN N.; SCHOERENBERGER H.; DAY, L., 2023. The State of Open Data 2023: A Digital Science. Digital Science, *Springer Nature*, Fligshare Eds. Disponible en: DOI [10.6084/m9.figshare.24428194.v1](https://doi.org/10.6084/m9.figshare.24428194.v1).

MA, X., 2022. Data Repository. En: SCHINTLER, L.; MCNEELY, C. (eds.), *Encyclopedia of Big Data*, 325-328. Cham: Springer International Publishing. ISBN: 978-3-319-32010-6. [10.1007/978-3-319-32010-6_59](https://doi.org/10.1007/978-3-319-32010-6_59).

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, 2023. *Estrategia Nacional de Ciencia Abierta 2023-2027*. [consulta diciembre 7, 2023]. Disponible en: <https://www.ciencia.gob.es/Estrategias-y-Planes/Estrategias/ENCA.html>

NUBE EUROPEA DE CIENCIA ABIERTA - EOSC, 2023. *Configurar el futuro digital de Europa*, 2023 [consulta diciembre 6, 2023]. Disponible en: <https://digital-strategy.ec.europa.eu/es/policies/open-science-cloud>

PAMPEL, Heinz et al., 2013. Making Research Data Repositories Visible: The re3data.org Registry. SULEMAN, Hussein (ed.), *PLoS ONE*. **8**(11), e78080. DOI [10.1371/journal.pone.0078080](https://doi.org/10.1371/journal.pone.0078080).

PARASTATIDIS, S., 2009. A Platform for All That We Know: Creating a KnowledgeDriven Research Infrastructure. En: HEY, T., et al., 2009.

The Fourth Paradigm: Data-Intensive Scientific Discovery, 165-172. Washington: Microsoft Research. ISBN: 978-0-9825442-0-4.
<https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/>

PARLAMENTO EUROPEO Y CONSEJO DE LA UNIÓN EUROPEA, 2016. Reglamento (UE) 2016/ 679 del Parlamento Europeo y del Consejo. Publicado el 27 de abril de 2016: relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la directiva 95/46/ ce (reglamento general de protección de datos). [consulta diciembre 7, 2023]. Disponible en: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>

UNESCO, 2023. *Recommendation on Open Science*, [consulta, diciembre 2023]. Disponible en: <https://www.unesco.org/en/open-science/about>

UNIVERSIDAD AUTÓNOMA DE MADRID, 2023. *E-ciencia datos*, [consulta, noviembre 17]. Disponible en: <https://www.uam.es/uam/vida-universitaria/bibliotecas/noticias/datasets>

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C., 2018. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436. [10.1016/j.jbusres.2017.12.043](https://doi.org/10.1016/j.jbusres.2017.12.043).