

CRÍSPULO TRAVIESO RODRÍGUEZ  
YOLANDA MARTÍN GONZÁLEZ  
(coords.)

# GESTIÓN Y REUTILIZACIÓN DE DATOS CIENTÍFICOS: FUNDAMENTOS Y EXPERIENCIAS



AQUILAFUENTE  
A



Ediciones Universidad  
**Salamanca**



Crípulo Travieso Rodríguez es doctor en Documentación y posee un Máster en Ciencia, Tecnología y Sociedad. Actualmente está vinculado como Profesor Titular al Departamento de Biblioteconomía y Documentación de la Universidad de Salamanca, donde imparte docencia en varios grados, como el Grado en Información y Documentación y el Grado en Lenguas, Literaturas y Culturas Románicas. Sus proyectos de investigación y publicaciones pertenecen al ámbito de la comunicación intercultural, la evaluación de la información científica y la organización de contenidos digitales.



Yolanda Martín González es Profesora Titular de Universidad del Departamento de Biblioteconomía y Documentación de la Universidad de Salamanca. Sus principales líneas de investigación se centran en la alfabetización informacional y en datos, así como en la gestión y reutilización de información científica y del sector público. Sobre estas temáticas ha publicado diversas monografías y artículos en editoriales y revistas de prestigio, además de haber dirigido proyectos financiados por entidades nacionales y europeas. Actualmente, imparte docencia en titulaciones oficiales de grado y postgrado en los ámbitos de la Información y la Documentación, y la Criminología.

GESTIÓN Y REUTILIZACIÓN DE DATOS  
CIENTÍFICOS: FUNDAMENTOS Y EXPERIENCIAS



CRÍSPULO TRAVIESO RODRÍGUEZ  
YOLANDA MARTÍN GONZÁLEZ  
(coords.)

GESTIÓN Y REUTILIZACIÓN  
DE DATOS CIENTÍFICOS:  
FUNDAMENTOS Y  
EXPERIENCIAS



Ediciones Universidad  
**Salamanca**

# AQUILAFUENTE, 386

© Ediciones Universidad de Salamanca y los autores

© Creación y composición de la cubierta: Intergraf

1ª edición: noviembre, 2025

ISBN: 978-84-1091-135-2 (pdf)

978-84-1091-136-9 (Impreso POD) / DL S 423-2025

DOI: <http://dx.doi.org/0AQ0386>

Ediciones Universidad de Salamanca

Plaza San Benito, 2

E-37002 Salamanca (España)

<http://www.eusal.es>

[eus@usal.es](mailto:eus@usal.es)

Composición y realización:

Intergraf

Impresión:

Nueva Graficesa

Av. de la Aldehuela, 80, 37003 Salamanca (España)

<https://www.graficesa.com/>

*Hecho en UE-Made in EU*

Obra evaluada por el sistema de doble ciego determinado por el Comité Científico del Congreso.



Usted es libre de compartir, copiar y redistribuir el material en cualquier medio o formato.

Ediciones Universidad de Salamanca no revocará mientras cumpla con los términos:

**ⓘ Reconocimiento** — Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.

**Ⓒ NoComercial** — No puede utilizar el material para una finalidad comercial.

**Ⓓ SinObraDerivada** — Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

Ediciones Universidad de Salamanca es miembro de la UNE

Unión de Editoriales Universitarias Españolas

[www.une.es](http://www.une.es)



Catalogación de editor en ONIX accesible en <https://www.dilve.es/>

## ÍNDICE

Presentación .....	9
Las bibliotecas universitarias en el ciclo de la comunicación científica .....	11
José Antonio MERLO VEGA	
Los <i>datasets</i> en abierto y su aportación al impacto social de la ciencia .....	33
Paloma G. BLÁZQUEZ, Almudena MANGAS-VEGA y Francisco José GARCÍA-PEÑALVO	
La curación de datos como metodología de investigación .....	51
Valeria MONTOYA-RONCANCIO e Inmaculada BRAVO GARCÍA	
Convergências entre a Folksonomia e a Curadoria Social e suas contribuições para a Ciência Cidadã .....	69
Gabriela DE OLIVEIRA SOUZA y Maria José VICENTINI JORENTE	
La importancia de los metadatos en la preservación de los registros de archivo digitales .....	89
Pedro Felipy CUNHA DA SILVA y Wagner JUNQUEIRA DE ARAÚJO	
Preservación digital y su vinculación con los datos abiertos de repositorios institucionales en España: un análisis descriptivo .....	105
Kevin LEÓN-GAVILANEZ	
Los servicios bibliotecarios de apoyo a la investigación en la Universidad de Salamanca .....	127
Almudena MANGAS-VEGA y Sonia SANTIAGO ROMÁN	
Interoperabilidad y acceso en los repositorios de datos de investigación de las universidades españolas .....	145
Ana B. RÍOS HILARIO y Carlos DÍAZ-REDONDO	
Autenticidad de los datos de investigación: análisis de modelos de metadatos en los repositorios españoles de RE3DATA .....	163
Marieta MARKS LÖW y Josefa GALLEGO LORENZO	
Un marco de métricas SMART para la mejora de los planes de gestión de datos .....	183
Federico GRASSO TORO	
As produções científicas brasileiras sobre a informação para a sustentabilidades na ciência a luz do projeto «estruturação da informação sobre a pesquisa brasileira» .	205
Ana Suely PINHO LOPES	

## INDEX

<i>Presentación</i> .....	9
<i>Academic Libraries in the Research Communication Cycle</i> ..... José Antonio MERLO VEGA	11
<i>Open Datasets and their Contribution to the Social Impact of Science</i> ..... Paloma G. BLÁZQUEZ, Almudena MANGAS-VEGA and Francisco José GARCÍA-PEÑALVO	33
<i>data Curation as Research Methodology</i> ..... Valeria MONTOYA-RONCANCIO and Inmaculada BRAVO GARCÍA	51
<i>Convergences between Folksonomy and Social Curation and their Contributions to Citizen Science</i> ..... Gabriela DE OLIVEIRA SOUZA and Maria José VICENTINI JORENTE	69
<i>The Importance of Metadata in the Preservation of Digital Archive Records</i> ..... Pedro Felipy CUNHA DA SILVA and Wagner JUNQUEIRA DE ARAÚJO	89
<i>Digital Preservation and its linkage to Open Data in Institutional Repositories in Spain: A Descriptive Analysis</i> ..... Kevin LEÓN-GAVILANEZ	105
<i>Research Support Library Services at the University of Salamanca</i> ..... Almudena MANGAS-VEGA and Sonia SANTIAGO ROMÁN	127
<i>Interoperability and Access in Spanish University Research Data Repositories</i> ..... Ana B. RÍOS HILARIO and Carlos DÍAZ-REDONDO	145
<i>Authenticity of Research Data: Analysis of Metadata Standards in Spanish Repositories at RE3DATA</i> ..... Marieta MARKS LÖW and Josefa GALLEGO LORENZO	163
<i>A SMART Metrics Framework for Improving Data Management Plans</i> ..... Federico GRASSO TORO	183
<i>Brazilian Scientific Productions on Information for Sustainability in Information Sciences in Light of the Project «Structuring Information On Brazilian Research»</i> ..... Ana Suely PINHO LOPES	205

## PRESENTACIÓN

La gestión y la reutilización de datos son pilares fundamentales de la ciencia abierta y promueven la transparencia, la colaboración ciudadana y la inclusión social.

En ese contexto, los datos de investigación se convierten en un elemento crucial, como unidades mínimas a partir de las que se construye y difunde el conocimiento científico. En la práctica, pueden tener múltiples vidas, ya que, si se gestionan adecuadamente, pueden ser reutilizados en diferentes contextos y por diferentes investigadores a lo largo del tiempo. De este modo, un dato puede servir para validar resultados de estudios previos, para explorar nuevas preguntas, para sustentar la toma de decisiones metodológicas o para ser combinado con otros conjuntos de datos en investigaciones interdisciplinarias. Además, con la extensión de los postulados de la ciencia abierta, los datos pueden ser compartidos globalmente, ampliando su impacto y utilidad en áreas que inicialmente no se habían considerado. Cada nueva utilización de los datos aumenta su valor, contribuyendo a un ciclo continuo de descubrimiento y avance del acervo científico.

Sin embargo, para que los datos de investigación sean accesibles, reutilizables e interoperables han de estar organizados según los estándares internacionales y acompañados de metadatos que describan su origen, calidad y condiciones de uso. Y para que esto sea efectivo es decisivo fomentar la alfabetización en datos tanto por parte de los docentes e investigadores como desde los servicios de apoyo a la investigación que proporcionan las bibliotecas universitarias, con el fin de formar a los investigadores no solo en el manejo técnico de los datos, sino también en su uso efectivo y crítico y en la capacidad para interpretar y reutilizar datos de manera responsable. Entran en juego, por tanto, diversos enfoques, configurando un objeto de estudio poliédrico, transversal y cambiante, en el que se hace necesario estar al tanto de las nuevas tendencias y enfoques para abordarlo.

El presente volumen está concebido, precisamente, dentro de esta posición plural y diversa a la hora de analizar los datos de investigación. El objetivo es conformar un repertorio de contribuciones que aborden distintos aspectos de esta unidad esencial de la comunicación científica. Como se comprobará en las siguientes páginas, la variedad de enfoques ofrece un panorama heterogéneo, pero también permite establecer algunas líneas comunes que parecen suscitar el máximo interés entre los profesionales de la información y la documentación. Entre ellas, además de la definición conceptual del propio dato de investigación, figuran cuestiones referidas a su descripción y a su preservación, únicas vías fiables para fomentar su posterior reutilización. También cabe mencionar, por su recurrencia, la labor de la biblioteca universitaria y de sus servicios de apoyo a la investigación como punto de referencia para los investigadores, tanto desde el punto de vista del tratamiento y búsqueda de datos como de la

formación en estos ámbitos, cada vez más demandada. La presencia de datos en los repositorios, institucionales o no, y su evaluación también aparecen de forma explícita y constante, así como las nuevas métricas y el impacto social que puede derivarse de su uso y difusión.

Confiamos en que esta muestra de intervenciones de especialistas de diversos ámbitos y orígenes sirva de aliciente para continuar abriendo interrogantes y sugiriendo ideas para futuras investigaciones, tanto en nuestro ámbito como en áreas adyacentes, y que las implicaciones científicas, sociales y culturales de la comunicación científica basada en datos contribuya a nuevos avances.

## LAS BIBLIOTECAS UNIVERSITARIAS EN EL CICLO DE LA COMUNICACIÓN CIENTÍFICA

### *ACADEMIC LIBRARIES IN THE RESEARCH COMMUNICATION CYCLE*

José Antonio MERLO VEGA

*Universidad de Salamanca. Departamento de Biblioteconomía y Documentación. España*  
merlo@usal.es

 <https://orcid.org/0000-0002-9102-4408>

**RESUMEN:** Las bibliotecas universitarias incluyen los servicios de apoyo a los procesos de la investigación como parte de su cartera de servicios. La investigación transcurre por diferentes fases desde que se diseña un proyecto hasta que se evalúa el resultado de las publicaciones producidas en dicha investigación. El estudio incluye un contexto teórico sobre los modelos de servicios bibliotecarios basados en ciclos de la investigación y aporta ejemplos de los más destacados. La metodología empleada se basa en la investigación bibliográfica y en el estudio de casos de bibliotecas universitarias de todo el mundo. Se analizan las diferentes fases del ciclo de la comunicación, que se estructuran a partir del modelo DIPE: Documentación, Investigación, Publicación y Expansión. Se describen las características de cada etapa y se desarrollan los procesos que intervienen en las mismas como base para la prestación de servicios bibliotecarios especializados y como identificación de las competencias que el personal bibliotecario debe poseer para prestar servicios de apoyo. Se establecen veinte procesos, cinco por cada etapa del modelo DIPE, explicándose los contenidos relacionados con cada proceso. Las modalidades de prestación de servicios de apoyo a la investigación se clasifican en cuatro grupos: formación, información, prestación y asistencia. Se describen las modalidades y sus derivadas, además de ofrecer una relación de los tipos de servicios bibliotecarios correspondientes a cada modalidad y de aportar ejemplos para cada uno de los cuarenta tipos de servicios identificados en el estudio. El capítulo ofrece una clasificación de modalidades y tipos de servicios de apoyo a la investigación y los enmarca en los procesos de la comunicación científica, como base para la planificación estructurada de los servicios que las bibliotecas pueden ofrecer al personal investigador de las universidades.

**PALABRAS CLAVE:** bibliotecas universitarias; apoyo a la investigación; comunicación científica; investigación; servicios bibliotecarios.

**ABSTRACT:** Academic libraries include support services for research processes as part of their portfolio of services. Research follows different phases from the design of a project to the evaluation of the outcome of the research published. A theoretical background on library service models based on research cycles is provided, giving examples of the most prominent ones. The methodology used is based on bibliographic research and case studies of university libraries around the world. This study analyses the different phases of the communication cycle and structures them based on the DIPE model: *Documentación* (Documentation), *Investigación* (Research), *Publicación* (Publication) and *Expansión* (Expansion). The characteristics of each stage are described, and the processes involved in each one are developed, as a basis for the provision of specialized library services and as an identification of the competencies that library staff must have in order to provide support services. Twenty processes are established, five for each stage of the DIPE model, and the contents related to each process are explained. The study classifies scientific communication support services into four categories: training, information, provision, and assistance. It describes the modalities and their derivatives, lists the types of library services corresponding to each modality and provides examples for each of the forty types of services identified in the study. The chapter provides a classification of modalities and types of research support services and frames them within the processes of scholarly communication, as a basis for structured planning of the services that libraries can offer to university research staff.

**KEYWORDS:** academic libraries; research support; scientific communication; research; library services.

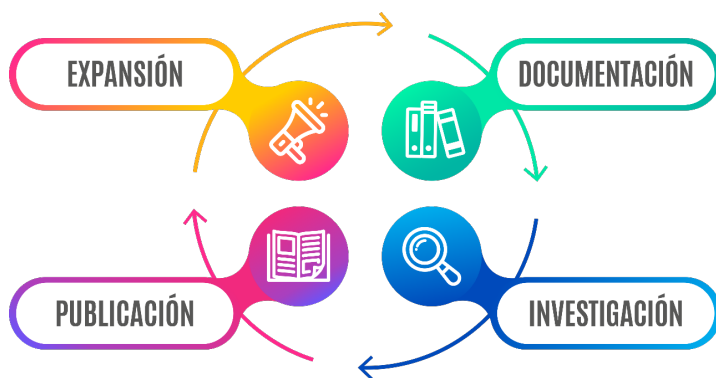
## 1. INTRODUCCIÓN

Las bibliotecas universitarias prestan servicios de apoyo a la investigación, enfocados a los diferentes procesos de la comunicación científica. La tipología de servicios es diversa, aunque se priorizan los relacionados con los sistemas de evaluación de la producción académica. Existe abundante bibliografía sobre la colaboración de las bibliotecas en los procesos de la investigación, de los que son ejemplos las aportaciones de Brown (2013), Atkinson (2016), Jubb (2016) como trabajos pioneros del enfoque actual. Más recientemente, Tran y Chan (2020) y Fernández y González (2021) analizan los tipos de servicios bibliotecarios orientados a la investigación y Merlo (2024) sistematiza los tipos de servicios especializados en investigación que las bibliotecas pueden ofrecer, que a menudo se circunscriben a los diferentes procesos de la ciencia. Ragon (2019) recopila modelos de ciclo de la investigación (*research lifecycle*) que se emplean en bibliotecas universitarias como marco para la prestación de servicios bibliotecarios, de la que es un ejemplo la Universidad de Utrech (Werner, 2017), que planificó los servicios bibliotecarios de investigación en tres grandes etapas (investigar, crear, compartir), definiendo en qué procesos específicos la biblioteca puede dar soporte a la investigación científica. Otros autores que relacionan servicios bibliotecarios y ciclo de la investigación son Chigwada (2020) y, especialmente, Hoffman (2016), quien distribuye el ciclo de vida de la investigación en etapas, asignando a cada una actividades de apoyo de las bibliotecas. Otros ejemplos de bibliotecas que emplean el ciclo de la comunicación científica para la presentación de servicios son

el modelo Research Lifecycle de la University of Central Florida (2021) y el de la Edinburgh Napier University (2024); en estos casos, los servicios se ajustan a las etapas y los procesos del ciclo de la comunicación científica. El ciclo de la comunicación científica se puede presentar de forma conjunta o complementaria con el ciclo de los datos de la investigación (Jiang *et al.*, 2025), como se puede comprobar en el modelo de la Universidad de Helsinki (Ojanen *et al.*, 2020).

La comunicación científica abarca todas las actividades que se realizan desde el diseño de una investigación hasta la evaluación de su impacto una vez publicada y comprende: la elaboración de un trabajo de investigación; la producción, difusión y evaluación de la investigación; y la preservación de las publicaciones (Maron *et al.*, 2019). Las bibliotecas prestan sus servicios de apoyo a la investigación en las diferentes etapas del ciclo de la comunicación. El objetivo de este capítulo es sistematizar las fases y las operaciones que se realizan en el desarrollo de una investigación científica y delimitar los servicios bibliotecarios especializados que se pueden prestar en cada una de las etapas y procesos. El siguiente apartado presenta el modelo DIPE (Figura 1), propuesto por Merlo (2022) y seguido por San-José *et al.* (2025), como medio de organización de los servicios bibliotecarios de investigación. El modelo DIPE se desarrolla de forma completa en el siguiente apartado de este capítulo y constituye un paradigma de las etapas y los procesos del ciclo de la comunicación científica y, a su vez, un marco referencial de servicios bibliotecarios de apoyo a la investigación, ya que dichos servicios se pueden orientar a las actividades desarrolladas en cada proceso. La estructuración del ciclo de la comunicación científica en las cuatro fases del modelo DIPE (Documentación, Investigación, Publicación, Expansión) es una excelente base para el diseño, la planificación y la prestación de servicios bibliotecarios especializados en investigación.

Figura 1. Modelo DIPE del ciclo de la comunicación científica



La metodología que se ha empleado para la elaboración del modelo DIPE se basó en dos técnicas de investigación clásicas: la revisión bibliográfica y el estudio de casos (Merlo, 2024). La revisión bibliográfica consistió en la obtención y el análisis de publicaciones sobre la implicación de las bibliotecas universitarias en los procesos de la comunicación científica. La

bibliografía se actualiza permanentemente y se puede consultar en Zotero, en el grupo público Biblioinvestigar, disponible en <https://www.zotero.org/groups/5536714/biblioinvestigar>.

El análisis de las publicaciones posibilitó realizar una clasificación de los servicios de apoyo a la investigación, que fue mejorada con el estudio de casos, en el que se procesaron un centenar de ejemplos de bibliotecas universitarias nacionales e internacionales que están prestando servicios de apoyo a la investigación. La sistematización de servicios de apoyo a la investigación que se presenta en este capítulo procede también de la experiencia del autor, docente en múltiples cursos en bibliotecas universitarias sobre el tema y director de un posgrado sobre servicios bibliotecarios especializados en investigación.

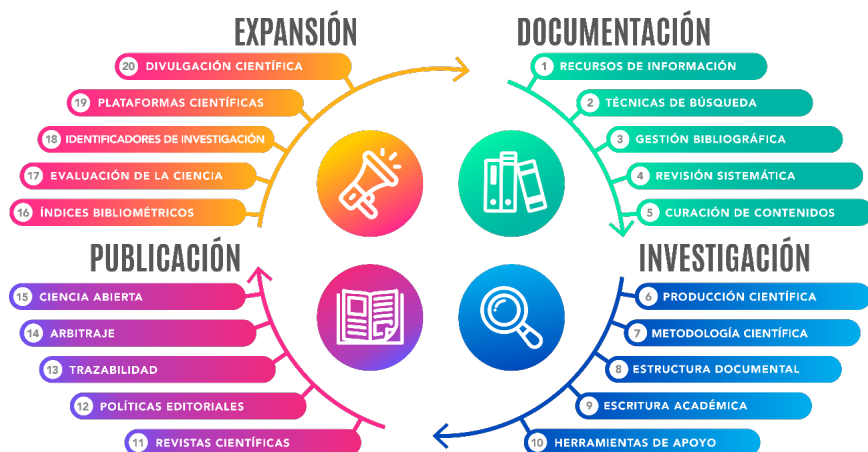
## 2. EL CICLO DE LA COMUNICACIÓN CIENTÍFICA

La comunicación científica es la suma de procesos secuenciales que intervienen en una investigación, que se inicia con la planificación de un tema de investigación y finaliza con la reutilización de la documentación generada y publicada como resultado de la investigación. La comunicación científica incluye procesos previos a la investigación propiamente dicha, como la recopilación de la bibliografía pertinente, así como actividades posteriores, tales como la divulgación y el análisis del impacto. El marco de la comunicación científica es más amplio que el de la investigación, por lo que es más conveniente denominar ciclo de la comunicación científica a todos los procesos de la investigación: diseño del estudio, investigación efectiva, elaboración y publicación de resultados de investigación y promoción y análisis de los datos y documentos producidos a partir de una investigación.

El ciclo de la investigación se puede estructurar en cuatro grandes fases (Figura 1): documentación (recursos bibliográficos y documentales), investigación (investigación propiamente dicha), publicación (resultados de la investigación) y expansión (difusión y evaluación). Las cuatro fases comprenden diferentes procesos consecutivos e interdependientes (Figura 2) que el personal de las bibliotecas debe conocer para poder prestar servicios de apoyo a la investigación. La fase de documentación integra los procesos de selección de los recursos documentales especializados, las técnicas de búsqueda avanzadas y expertas, los sistemas de gestión bibliográfica de la información, la revisión sistemática y los servicios de curación de contenidos; procesos que permitirán el triaje documental, es decir, la selección de la información de interés para el tema de estudio. La fase de investigación se refiere a la investigación propiamente dicha, comprendiendo la variedad de tipologías documentales con las que se comunica la ciencia, las metodologías científicas, la estructura de un texto científico, las características de la escritura académica y las herramientas que servirán de apoyo a la investigación. La fase de publicación integra los procesos de identificación de las revistas más convenientes en función de sus políticas, temáticas e impacto; la trazabilidad del proceso de publicación, los sistemas de arbitraje, y los canales de difusión en acceso abierto. La fase final se denomina expansión, por su doble carácter de difusión y evaluación; agrupa los procesos del análisis de impacto bibliométrico y de evaluación de la ciencia, recursos de promoción, identificadores y perfiles académicos, las plataformas científicas y los medios de divulgación científica.

Los siguientes subapartados desarrollan los veinte procesos que se realizan en las cuatro fases del modelo DIPE del ciclo de la comunicación científica.

Figura 2. Procesos de la comunicación científica en el Modelo DIPE



## 2.1. DOCUMENTACIÓN

Fase de planificación, en la que se diseña la investigación y se seleccionan los recursos bibliográficos, documentales e instrumentales necesarios. Todas las investigaciones deben iniciarse localizando las publicaciones previas que existan sobre el mismo tema o aspectos relacionados. Los procesos que intervienen en esta fase comprenden la selección de recursos de información; el uso avanzado de técnicas de recuperación de información en fuentes especializadas, la organización en gestores bibliográficos de la información y de los documentos recuperados; la aplicación de métodos de revisión sistemática de la producción científica, y el diseño de sistemas de actualización, triaje documental y curación de contenidos.

Las competencias del personal bibliotecario que ofrezca servicios de apoyo en esta fase de la comunicación científica se corresponden con conocimientos y habilidades relativas a las tipologías de recursos de información de valor para la investigación, técnicas avanzadas de recuperación de información, procedimientos de gestión bibliográfica, modelos de revisión bibliográfica y procedimientos de curación de contenidos.

1. Recursos de información. Conocer las fuentes de información disponibles es el primer paso para realizar una investigación. Las publicaciones de potencial interés pueden ser localizadas mediante herramientas bibliotecarias, buscadores académicos, plataformas editoriales, repositorios y recolectores y en localizadores de documentos. Las herramientas bibliotecarias son los productos que las bibliotecas producen para la recuperación de información, como

catálogos, sistemas de descubrimiento o guías informativas. Los buscadores académicos son plataformas de contenidos científicos que se pueden consultar gratuitamente, entre los que se encuentran los buscadores científicos, los agregadores científicos, las redes científicas y las plataformas de revistas en abierto. Las plataformas editoriales son los portales y los recursos de las editoriales y las empresas especializadas en publicaciones científicas, que incluyen las categorías de bases de datos multidisciplinares y especializadas, portales académicos y plataformas comerciales. Los repositorios y los recolectores facilitan documentos de carácter científico archivados para su acceso abierto. Por último, los localizadores de documentos son servicios que buscan versiones abiertas o autorizadas de publicaciones, ya sea mediante aplicaciones, extensiones de navegadores o servicios web de localización de documentos.

2. Técnicas de búsqueda. Saber emplear técnicas de búsqueda avanzadas y expertas permitirá realizar consultas precisas y obtener resultados pertinentes. Las fuentes de información disponen de herramientas de consulta, que incluyen búsquedas expertas, tesauros, filtros, etc. No es suficiente con el conocimiento de los recursos de información de interés sobre un tema de investigación, sino que es necesario dominar las diferentes posibilidades que ofrece cada fuente de información para buscar de forma precisa y, por consiguiente, recuperar solo resultados relevantes. El uso de búsquedas avanzadas, combinando campos, pero, especialmente, la recuperación experta mediante comandos específicos y términos de tesoro, permitirán localizar de forma rápida publicaciones relevantes. Asimismo, el correcto manejo de los buscadores basados en inteligencia artificial se incluye en esta etapa de la investigación.

3. Gestión bibliográfica. Trabajar con herramientas de gestión bibliográfica permitirá reutilizar la documentación, consultarla y mencionarla. Las aplicaciones o servicios web de gestión bibliográfica son sistemas integrales indispensables para la investigación, ya que son herramientas para la organización de documentos, tanto información bibliográfica como recursos a texto completo. Entre sus características destacan que las referencias pueden ser compartidas con otros usuarios interesados en el mismo tema y que permiten trabajar con las referencias desde un equipo local y/o a través de Internet, sincronizando los registros. Los gestores bibliográficos ofrecen soluciones para importar referencias procedentes de catálogos bibliográficos, bases de datos, otras bibliografías, etc., y se integran con procesadores de texto; igualmente, exportan las referencias en diferentes codificaciones. Su principal funcionalidad es la posibilidad de crear bibliografías siguiendo múltiples estilos; además, permiten identificar publicaciones, organizarlas en carpetas, leerlas y anotarlas, asignar materias a cada documento, compartir en grupos, recibir alertas, revisar duplicados, etc.

4. Revisión sistemática. Establecer sistemas de revisión sistemática facilitará conocer el estado de la cuestión de un tema de investigación y disponer de recursos de información actualizados. El concepto de revisión sistemática oscila entre, por una parte, los procedimientos normalizados para la recuperación de publicaciones recientes sobre un tema de investigación, para disponer de la bibliografía actualizada y, por otra parte, los tipos específicos de publicaciones consistentes en que el método y el resultado es la propia revisión de literatura profesional.

La localización y el análisis de las publicaciones previas acerca de un tema sobre el que se está investigando pueden considerarse revisión sistemática si se hacen mediante procedimientos normalizados, aunque en este caso es más preciso denominar a este proceso revisión bibliográfica (*literature review*); se trataría de las consultas que se diseñan y se repiten a partir de estrategias definidas, para establecer el estado del arte o contexto teórico de un tema de investigación, que se empleará como bibliografía en la publicación resultante de la investigación para la que han localizado las publicaciones. Una revisión bibliográfica/sistemática es también un tipo de publicación, para cuya elaboración se ha empleado una metodología específica de realización de revisiones, que incluye una descripción narrativa del proceso (síntesis) y una exposición de los resultados de la revisión (análisis); se trata de un documento en sí mismo (*paper, review, systematic review*).

5. Curación de contenidos. Emplear técnicas de descubrimiento de nuevas publicaciones permite recibir de forma inmediata información de novedades bibliográficas, para su posterior revisión y uso. El proceso de actualización puede realizarse de forma automática mediante técnicas y sistemas de curación de contenidos, gracias a servicios de sindicación o sistemas de selección de contenidos para su recepción automática, medios de agregación para la recepción y lectura de forma unificada de información de interés y plataformas de descubrimiento desde las que se reciben alertas selectivas, por relaciones en redes sociales o por perfiles temáticos. Es especialmente útil la tecnología de sindicación de contenidos que emplean las publicaciones digitales para poder ser consultadas desde servicios o herramientas externas. Las revistas electrónicas, los blogs, las webs y los medios informativos permiten recibir (sindicar) sus publicaciones para agregarlas en servicios o herramientas de lectura/agregación de contenidos. Las aplicaciones de curación de contenidos son parte de la estrategia de triaje o selección documental que se debe diseñar para disponer de toda la información bibliográfica necesaria para la investigación. El triaje incluye: curación de contenidos para agregar publicaciones y posteriormente revisarlas, selección de las que se consideren útiles y gestión bibliográfica para archivar la información utilizable.

## 2.2. INVESTIGACIÓN

Fase de producción, en la que se realiza la investigación activa y se aplican metodologías para la obtención de resultados que permitan confirmar hipótesis. La investigación propiamente dicha se realiza en esta segunda fase de la comunicación científica. El tipo de publicación derivada de la investigación requerirá características específicas, tanto de contenido como de estructura: no es lo mismo investigar para publicar un artículo científico que para presentar una ponencia en un congreso o para defender una tesis doctoral. Los procesos que se incluyen en esta fase comienzan con la identificación de los tipos de documentos en los que se materializa la producción científica; siguen con la selección de métodos de investigación y la aplicación de las técnicas correspondientes, y finalizan con los conocimientos sobre estructura de una publicación, características de la escritura académica y oferta tecnológica de apoyo a la investigación.

Las competencias que se demandan del personal bibliotecario en la fase de Investigación se centrarán, por tanto, en la metodología empleada en la investigación científica, tipología y estructura de los documentos en los que se difunden los resultados de la investigación, así como en la redacción científica y en las herramientas tecnológicas de desarrollo de investigaciones.

6. Producción científica. Conocer las formas de la producción científica es necesario para saber cómo se concretarán los resultados de una investigación. La producción científica se materializa en diferentes tipos de publicaciones, sobre todo textuales. El principal producto de difusión de los resultados de la investigación son los artículos científicos, que se publican en las revistas científicas. Además de artículos científicos, las publicaciones de cualquier disciplina pueden tener forma de libros, capítulos, comunicaciones y ponencias, tesis y disertaciones, recursos de internet, documentos técnicos y recursos multimedia. Diferentes disciplinas emplean tipologías documentales específicas, como test en educación, ensayos clínicos o protocolos en medicina y estudios de caso en economía, por ejemplo. El artículo científico (*paper*) es el principal tipo de documento en el que se difunden los resultados de las investigaciones, mientras que la revista científica (*journal*) es la vía prioritaria de publicación y difusión de la investigación. Los tipos de artículos varían en función de sus objetivos y estructuras. Las revistas científicas indican qué tipo de artículos publican y qué características deben cumplir. La tipología de artículos que se publican en revistas científicas suele incluir artículos de investigación (aplicación del método científico), artículo técnico (descripción de productos o servicios), estudios de caso (descripción de prácticas destacadas), revisión bibliográfica (análisis de publicaciones), artículos de opinión (puntos de vista subjetivos) y ensayos (exposición de temas).

7. Metodología científica. Saber aplicar una metodología científica conveniente será la base de cualquier investigación científica. La ciencia emplea diferentes métodos de investigación para demostrar o rechazar hipótesis. Cada método utiliza las técnicas de investigación más apropiadas para obtener los resultados sobre los que se apoyen las conclusiones. La investigación debe cumplir determinados principios científicos: plantear cuestiones que puedan ser demostradas empíricamente, conectar la investigación con la teoría conocida, utilizar métodos que permitan la investigación directa del tema de estudio, relacionar los argumentos de forma coherente, replicar métodos anteriores o ser replicable en estudios posteriores, así como publicar los resultados para la difusión y el análisis por pares científicos. Los métodos de investigación se dividen en cuantitativos (datos) y cualitativos (percepciones), aunque también hay métodos mixtos. Una investigación se diseña en fases: planteamiento del problema que se quiere estudiar, síntesis del contexto teórico, definición de objetivos, aplicación de la metodología pertinente, análisis de resultados y comprobación de la hipótesis. La selección de las técnicas de investigación es un aspecto esencial para el correcto desarrollo de la investigación. Existe una amplia variedad de técnicas de investigación, como, por ejemplo: estudio de caso, análisis bibliométrico, experimento, entrevistas, cuestionarios, revisiones bibliográficas, fuentes históricas, observación, análisis de registros informáticos, etc.

8. Estructura documental. Conocer las convenciones acerca de la estructura de las publicaciones permitirá presentar los resultados de la investigación de forma académica. Los artículos científicos siguen una estructura documental similar para exponer de forma clara hipótesis, metodologías, resultados y conclusiones. Los artículos contienen una parte de identificación común compuesta de título, resumen y palabras clave, tanto en el idioma del artículo como en inglés, como lengua estándar. Los contenidos siguen una división comúnmente aceptada compuesta de introducción (antecedentes, hipótesis, contexto teórico), metodología (recursos, métodos y técnica de investigación), resultados (datos y hallazgos obtenidos), discusión y conclusiones (comprobación de objetivos y aportaciones) y bibliografía (publicaciones empleadas en el estudio). Habitualmente, las normativas de las revistas científicas incluyen pautas para estructurar los artículos en función de la tipología de la investigación.

9. Escritura académica. Emplear las normas de la escritura académica diferenciará una publicación científica de un documento informativo, literario o de divulgación. El estilo de redacción de textos en la comunicación científica se orienta hacia la transmisión de los resultados de la investigación de forma descriptiva, siguiendo pautas de escritura académica que evitan el lenguaje retórico mediante una redacción expresiva, fluida y precisa. Las reglas ortográficas y gramaticales son la base de la escritura y deben ser respetadas en la escritura científica. Los manuales de estilo ayudan a resolver dudas y a solucionar problemas. Igualmente, las editoriales indican también cómo debe presentarse un texto desde el punto de vista ortotipográfico. El idioma inglés se considera el idioma de la ciencia, por lo que el inglés neutro se usa en todo el mundo para la redacción de textos académicos, aplicando los mismos criterios de claridad, precisión y brevedad.

10. Herramientas de apoyo. Conocer y emplear la oferta específica de recursos de apoyo a la investigación permitirá la realización ágil de estudios científicos. La investigación emplea diversos tipos de herramientas de apoyo en sus diferentes etapas, sobre todo de tipo tecnológico. Recursos de apoyo de uso genérico son plataformas y servicios de encuestas, mapas conceptuales, planificación de proyectos, gestión bibliográfica, almacenamiento remoto o redacción colaborativa. Además, hay programas de uso específico para la investigación cuantitativa y la investigación cualitativa, que a menudo incluyen soluciones de código abierto.

### 2.3. PUBLICACIÓN

Fase de publicación, en la que se realiza la comunicación científica en sentido estricto, como etapa en la que se inicia la difusión de los resultados de la investigación. El objetivo de la investigación es la comunicación pública de sus resultados, es decir, la publicación de los hallazgos resultantes de la aplicación de una metodología científica para la demostración de una hipótesis o para la exposición analítica de los hechos estudiados. La publicación de los resultados de la investigación se realiza mayoritariamente en revistas científicas, de las que hay que conocer sus políticas editoriales y sus procesos de gestión y arbitraje. Igualmente, la publicación de productos

derivados de la investigación se realiza en acceso abierto, ya sea en repositorios, con el matiz de poner en público lo ya publicado en revistas, así como en revistas científicas electrónicas, que publican en abierto con los mismos criterios que cualquier revista de carácter académico.

Las competencias, entendidas siempre como la suma de conocimientos y habilidades que debe poseer el personal bibliotecario encargado del apoyo a la investigación en esta fase de Publicación, deben versar sobre edición académica, características de las revistas científicas, procesos de publicación y revisión, así como acerca de repositorios como medios de publicación y difusión de los resultados de la investigación.

11. Revistas científicas. Los resultados de la investigación se publican en revistas científicas, por lo que hay que conocer qué publicaciones son acordes con el tema del análisis realizado. Las revistas científicas son la principal vía de difusión de los resultados de la investigación, se identifican mediante códigos normalizados, siguen políticas editoriales y son evaluadas en función de su impacto. Se trata de publicaciones periódicas, editadas por empresas o instituciones, sujetas a estándares, que siguen políticas editoriales, que publican contenidos en forma de artículos, que se difunden de forma comercial o abierta y que se integran en bases de datos y agregadores científicos. Las revistas científicas pueden publicar investigaciones inéditas (*journals*), incluir experiencias profesionales (*newsletters*) o enfocarse en la divulgación científica (*magazines*). Las revistas científicas publican investigaciones originales, siguiendo un método científico, como forma de transmisión del conocimiento. Se trata de contenidos especializados que interesan a comunidades científicas determinadas.

12. Políticas editoriales. Las revistas siguen sus propias políticas editoriales que es necesario conocer para decidir en cuál publicar y cómo se gestionarán los derechos. Las directrices o políticas de una revista científica pueden ser propias, de la editorial o la revista, o acordadas por consorcios de editoriales académicas para normalizar sus procesos y para establecer buenas prácticas y pautas de comportamiento ético. Las políticas editoriales deben determinar su alcance y definir aspectos como periodicidad, idiomas, tipología de contribuciones, estructura de los artículos, sistemas de arbitraje, detección de plagios, conductas éticas, costos de publicación, licencias de acceso abierto y preservación a largo plazo. Asimismo, las revistas deben configurar comités editoriales (redacción, asesoramiento, por ejemplo), así como el seguimiento de directrices internacionales: normalización, identificadores, metadatos, autoridades, políticas de ética, transparencia en la evaluación, trazabilidad de los procesos.

13. Trazabilidad. Las investigaciones enviadas a revistas para su publicación siguen un proceso de trazabilidad que permite conocer el estado de la propuesta de publicación. La publicación de resultados de investigación sigue un proceso estructurado cuya trazabilidad permite conocer los elementos que intervienen en cada fase. El proceso de publicación incluye el envío de propuestas para publicar, la gestión de originales, la evaluación, la revisión y la publicación final. Las propuestas de originales (manuscritos, *preprints*) se envían siguiendo los procedimientos que indiquen las revistas. Las revistas ofrecen información para autores/as donde explican el procedimiento. El proceso requiere el alta en la plataforma para poder realizar un seguimiento

de todas las etapas (trazabilidad). Las revistas ofrecen información sobre sus normas y sobre el registro en la plataforma que empleen. La trazabilidad puede estructurarse en cinco fases: envío, recepción, evaluación, comunicación y publicación. La fase de comunicación incluye la posibilidad de revisión y mejora a partir de los resultados de la evaluación. Muchas empresas relacionadas con la edición y la comunicación científica están ofreciendo servicios de apoyo a la publicación, que suelen incluir servicios como revisión técnica, traducción, prearbitraje, selección de publicación, gestión de envíos a revistas, intermediación entre revistas y autores, gestión de tasas de publicación, difusión de resultados, etc.

14. Arbitraje. Las revistas científicas siguen modelos de arbitraje para evaluar las propuestas de publicación y decidir si las aceptan, modifican o rechazan. La revisión de las propuestas de publicación se realiza mediante sistemas de arbitraje destinados a que evaluadores externos valoren la calidad de los originales y su adecuación a las políticas editoriales de la revista. Los métodos de evaluación o arbitraje que emplean las revistas son revisión por pares anónima única, revisión por pares doblemente anónima (doble ciego), revisión por pares abierta, revisión por pares posterior a la publicación y elaboración de informes por especialistas. Se emplea el concepto de pares para designar a evaluadores conocedores de la materia del artículo. El anonimato (ciego) se refiere a que las personas evaluadas y las evaluadoras no se identifican en el proceso de revisión. El método de evaluación doble ciego sigue siendo el más empleado en las revisiones de originales. Igualmente, la revisión abierta, consistente en la evaluación posterior a la publicación provisional, realizada de forma pública, también es un modelo en crecimiento.

15. Ciencia abierta. La difusión de investigaciones siguiendo principios de ciencia abierta posibilita que las publicaciones se puedan difundir de forma universal. Ciencia abierta es la suma de los procesos de la comunicación científica que se realizan de forma pública. Se deben entender como publicaciones en acceso abierto aquellas que, además de posibilitar su lectura, copia y distribución, siguen los protocolos que permiten el intercambio de datos, la integración en repositorios y la localización por medio de recolectores o motores de búsqueda especializados en archivos abiertos. La difusión de los resultados de la investigación se realiza también en repositorios institucionales, tanto versiones publicadas como inéditas. La publicación de los resultados de la investigación en repositorios (ruta verde) implica una serie de características: los documentos se depositan en formatos digitales, los artículos pueden estar previamente publicados o ser inéditos, el documento está accesible de forma íntegra, los datos son interoperables con recursos externos de recolección o agregación, por lo que la información depositada es ampliamente visible. Por otra parte, la publicación en revistas en acceso abierto viene determinada por las políticas editoriales, abriendo el acceso a todos los contenidos de forma gratuita para autores y lectores (ruta diamante) o estableciendo límites al acceso (ruta dorada). Las revistas de la ruta dorada y diamante siguen los mismos procesos que cualquier revista científica en cuanto a políticas, normas de publicación, arbitraje y trazabilidad.

## 2.4. EXPANSIÓN

Fase de promoción de la investigación. La investigación publicada debe ser promocionada mediante la divulgación de los productos documentales derivados de la misma, ya sean artículos, datos de investigación o cualquier otra tipología documental de la producción científica; asimismo, se deberá analizar el impacto de las publicaciones, entendido como la presencia de una investigación en otras publicaciones, ya sea empleada como contexto o mención en posteriores publicaciones. Igualmente, la visibilidad de una investigación publicada amplía también el impacto y el uso del conjunto de la producción científica de una persona, grupo de investigación, departamento universitario, institución de enseñanza superior o instituto de investigación. Además del impacto de las publicaciones, en esta fase del ciclo de la comunicación científica intervienen otros elementos de evaluación de la ciencia, sistemas de identificación de obras, personas y entidades, producción de plataformas y de perfiles de autores y universidades y medios de divulgación científica.

El personal bibliotecario con responsabilidad en los servicios de apoyo a la investigación debe tener competencias sobre medios de comunicación y divulgación científica, procedimientos de evaluación de la ciencia e indicadores bibliométricos que se emplean para analizar el impacto de las publicaciones científicas.

16. Índices bibliométricos. El impacto de las publicaciones científicas se obtiene a partir de los índices bibliométricos, que se emplean para la evaluación de la investigación. La bibliometría mide el impacto de las publicaciones mediante indicadores que tienen en cuenta el número de artículos publicados y el número de citas recibidas. Los indicadores bibliométricos de referencia se emplean en todo el mundo como el principal método de evaluación de la ciencia. Se trata de índices con datos cuantitativos que miden el impacto de la producción científica. El impacto se simplifica mediante la aplicación de fórmulas que tienen en cuenta esencialmente el número de publicaciones y el número de citas recibidas en un rango temporal determinado. Los indicadores bibliométricos se obtienen a partir de las publicaciones de bases de datos concretas, por lo que sus datos están limitados a la producción previamente seleccionada en las fuentes que hayan decidido las empresas productoras de los índices de impacto.

17. Evaluación de la ciencia. Conocer los diferentes sistemas y políticas de evaluación de la ciencia servirá para orientar el tipo de investigación y los medios de publicación. La evaluación de la ciencia se realiza teniendo en cuenta el factor de impacto y la calidad de las revistas científicas, para lo que se emplean diferentes métodos de análisis. Una de las formas complementarias de evaluación de la producción científica son las métricas alternativas, que ofrecen indicadores de uso de los recursos de información. Las métricas alternativas (*altmetrics*) son especialmente útiles porque miden el impacto de los documentos, no de las revistas donde se han publicado, teniendo en cuenta aspectos que van más allá de las citas. La evaluación de la ciencia también incluye a los *rankings* académicos, que tratan de ofrecer información comparativa de universidades y centros de investigación.

18. Identificadores de investigación. La difusión de la producción científica se realiza por diferentes vías, entre las que se encuentra la identificación inequívoca de las publicaciones procedentes de una institución. La autoría de las investigaciones se controla mediante identificadores de investigación, que permiten determinar de forma precisa la responsabilidad de las publicaciones. El control de la producción científica requiere la identificación de autores, publicaciones e instituciones. La difusión de la producción científica necesita que se identifique claramente la responsabilidad intelectual y la entidad en la que se trabaja. Los identificadores permanentes son sistemas de codificación de las formas de autor, institución y publicación para su uso en recursos bibliográficos y bibliométricos, así como para la gestión de la investigación. La normalización evita problemas de ambigüedad y de asignación incorrecta de trabajos a personas no autoras de los mismos. Además de identificar autorías y documentos, las instituciones elaboran pautas para la firma científica, para asegurarse de que toda la producción científica es recogida de forma adecuada en los grandes índices bibliométricos y *rankings* científicos.

19. Plataformas científicas. Las instituciones difunden la producción de sus centros mediante plataformas científicas en las que se reúne la investigación realizada. Además de mediante el uso de identificadores, la promoción de las publicaciones se realiza a través de portales de perfiles de autores, redes académicas y plataformas de producción científica. Los perfiles académicos son las páginas que muchas plataformas dedican a cada investigador, con sus datos institucionales, identificadores, indicadores bibliométricos y publicaciones; pueden ser gestionados por los propios investigadores o bien solo por la organización propietaria de la plataforma o sus entidades colaboradoras. Las redes académicas permiten difundir la producción científica a través de plataformas públicas de gestión privada y fines comerciales, aunque tienen la ventaja de que permiten dar a conocer las publicaciones y, por tanto, aumentar el número de citas que se recibirán; otra ventaja es la cooperación entre investigadores de los mismos campos temáticos, ya que permiten realizar seguimientos, establecer grupos y solicitar documentos. Las plataformas de producción científica son los portales desde donde las instituciones difunden su producción científica. Las universidades y las entidades de investigación emplean sistemas de gestión y difusión de sus publicaciones. Por una parte, se encuentran los sistemas de gestión curricular de la docencia y la investigación (*Current research information system*, CRIS), que no son medios de difusión, sino de administración de la investigación; pero, por otro lado, los centros de investigación disponen de portales de producción científica, que son las plataformas desde las que entidades de investigación reúnen y difunden su producción científica, con indicadores y acceso abierto a contenidos propios o de consorcios a los que se pertenece.

20. Divulgación científica. La investigación debe comunicarse mediante técnicas de divulgación científica que permitan que la producción científica sea conocida y compartida. La difusión de la investigación debe incluir también la divulgación de las publicaciones y los resultados de la investigación, realizada mediante técnicas periodísticas y de comunicación. La divulgación científica se efectúa con medios de comunicación personales, institucionales o comerciales. La divulgación consiste en la redacción de noticias relacionadas con resultados

de la investigación, que se difunden en las agencias de noticias científicas, los servicios de comunicación de las entidades, en las redes sociales personales e institucionales y en los canales informativos de las revistas o los organismos en los que trabajen los autores de las investigaciones. Igualmente, la difusión de la investigación se realiza mediante su presentación en congresos y jornadas científicas, en las que se exponen bien los avances de una investigación en curso, bien los resultados de la misma, una vez finalizada la investigación.

### 3. SERVICIOS DE APOYO A LA INVESTIGACIÓN

Las bibliotecas de universidades e instituciones científicas han incluido los servicios de apoyo a la investigación dentro de su cartera de servicios. La bibliografía profesional sobre bibliotecas y apoyo a la investigación es abundante, aunque apenas se han publicado estudios como los de Rey Martín *et al.* (2018) o informes como el de la Comisión Asesora de Bibliotecas y Documentación de Chile (2020), en los que se ofrece una tipología de servicios prestados o potencialmente prestables en las bibliotecas universitarias de apoyo a los procesos propios de la investigación. Igualmente, el 5.º Plan Estratégico de REBIUN, correspondiente al periodo 2024-2027, incluye un ámbito de servicios de apoyo a la investigación y establece áreas de actuación (Rebiun, 2024): acreditación y sexenios, impacto de las publicaciones, ética en la publicación, colaboración científica, gestión de datos de investigación, gestión de publicaciones en repositorios institucionales, interoperabilidad con CRIS y portales de investigación, propiedad intelectual, visibilidad de los servicios, humanidades digitales, requisitos de los repositorios institucionales.

Merlo (2024) considera que los servicios que las bibliotecas ofrecen van más allá de la investigación propiamente dicha y aboga por emplear la denominación servicios bibliotecarios especializados en comunicación científica, que se definen como «las prestaciones que las bibliotecas ofrecen para ayudar al usuario en los procesos de una investigación: selección y gestión de recursos documentales y bibliográficos, aplicación de metodologías científicas, publicación de resultados de la investigación y difusión y evaluación del alcance de la producción científica». Igualmente, este autor realiza un análisis de los tipos de servicios que las bibliotecas ofrecen y aporta una clasificación de modalidades y un catálogo de servicios de apoyo de las bibliotecas en todos los procesos del ciclo de la comunicación científica. Este capítulo emplea la clasificación de servicios difundida en dicha publicación, con la intención de ofrecer un estudio completo en el que se contemplen tanto los procesos del ciclo de la comunicación científica establecidos en el modelo DIPE como las modalidades de servicios y las tipologías de los mismos.

Los tipos de servicios que las bibliotecas ofrecen para apoyar a sus universidades y al personal investigador en todos los procesos del ciclo de la comunicación científica pueden clasificarse en cuatro grandes modalidades: formación, información, prestación y asistencia (Tabla 1). Los servicios de formación comprenden todas las actividades de alfabetización informacional relacionadas con los procesos y los recursos de la investigación: cursos, tutoriales y actividades instructivas para la adquisición de competencias en comunicación científica a

través de los conocimientos y las habilidades específicos según los tipos de investigación. Los servicios informativos se orientan a la producción de recursos informativos sobre investigación: boletines, guías temáticas, productos informativos, medios y servicios de comunicación que informen sobre recursos documentales, técnicas de investigación, publicación de resultados y difusión de la producción científica. Los servicios de prestación se centran en la elaboración de productos bibliográficos y documentales para la investigación: búsquedas, informes y estudios, información elaborada y servicios requeridos en los procesos de la comunicación científica y la evaluación de la ciencia. Por último, los servicios de asistencia se orientan hacia la colaboración en el desarrollo de la investigación: servicios de consultoría, apoyo funcional a la gestión de la investigación y la administración de plataformas, asistencia en relación con infraestructuras, equipamientos, aplicaciones, normativas y políticas de investigación.

Tabla 1. Modalidades de servicios bibliotecarios especializados en investigación

<b>A. FORMACIÓN</b>	<b>C. PRESTACIÓN</b>
A1. Talleres formativos	C1. Referencia especializada
A2. Materiales instructivos	C2. Informes
<b>B. INFORMACIÓN</b>	C3. Equipo de trabajo
B1. Guías web	C4. Gestión técnica
B2. Guías informativas	<b>D. ASISTENCIA</b>
B3. Servicios de comunicación	D1. Atención a la investigación
B4. Canales informativos	D2. Soporte informativo

Los distintos tipos de servicios correspondientes a cada modalidad se detallan en los siguientes subapartados. En cada caso, se describe la modalidad de servicios de apoyo y los tipos concretos de servicios bibliotecarios que se enmarcan en cada modalidad. Además, se ofrece un ejemplo de servicio concreto para cada una de las cuarenta tipologías identificadas. La clasificación que se desarrolla en los próximos párrafos es una revisión, actualización y reorganización de la publicada en el artículo «Servicios bibliotecarios especializados en investigación: competencias, modalidades y catálogo de servicios» (Merlo, 2024), que se considera complementario a este capítulo. El desarrollo de las modalidades y servicios se realiza partiendo de las cuatro modalidades (A-D), de sus correspondientes doce submodalidades (A1, A2, etc.) y de las cuarenta tipologías de los servicios (1-40), que se describen dentro de sus respectivas modalidades y submodalidades.

## A. SERVICIOS DE FORMACIÓN

La biblioteca ofrece a los usuarios servicios de alfabetización informacional especializados en comunicación científica. La formación se realiza de forma directa, mediante cursos, talleres y seminarios (A1 Talleres formativos); o indirecta, a través de materiales formativos (A2 Tutoriales).

### A1. *Talleres formativos*

1. Talleres formativos. Seminarios sobre cualquier proceso, operación o recurso del ciclo de la comunicación científica. Ejemplo: Taller de búsquedas en bases de datos especializadas.

### A2. *Materiales instructivos*

2. Tutoriales. Recursos instructivos relacionados con investigación y comunicación científica. Ejemplo: Tutorial sobre el uso de Zotero.

## B. SERVICIOS DE INFORMACIÓN

La biblioteca elabora productos informativos para los usuarios sobre aspectos relacionados con la investigación. Los productos informativos relacionados con los procesos de la comunicación científica son servicios de apoyo a la investigación. La biblioteca puede realizar información vía web (B1 Guías web); redactar boletines o folletos (B2 Guías informativas); crear listas informativas (B3 Servicios de comunicación); y mantener redes o medios sociales (B4 Canales informativos).

### B1. *Guías web*

3. Biblioguías. Información web estructurada con contenidos textuales, multimedia y recursos de utilidad para la elaboración de trabajos de investigación. Ejemplo: Biblioguía (*libguide*) de recursos de investigación en Trabajo Social.

### B2. *Guías informativas*

4. Guías informativas. Folletos, infografías, boletines y documentos elaborados sobre temas de investigación. Ejemplo: Infografía sobre licencias de uso de artículos científicos.

### B3. *Servicios de comunicación*

5. Servicios de comunicación. Listas de distribución o correo de carácter informativo. Ejemplos: Lista de distribución sobre recursos para la investigación.

#### B4. *Canales informativos*

6. Canales de mensajería. Difusión de información mediante canales de mensajería y redes sociales. Ejemplo: Canal en aplicación móvil para difundir información sobre investigación.

#### C. SERVICIOS DE PRESTACIÓN

Los usuarios demandan a la biblioteca informaciones o documentos para la producción o evaluación de investigaciones; servicios bajo demanda, en los que la biblioteca entrega productos o informaciones previa petición individual o institucional. Los servicios de este tipo pueden atender consultas (C1 Referencia especializada); redactar estudios de situación (C2 Informes); prestar servicios integrados en grupos de investigación (C3 Equipo de trabajo); y administrar plataformas, aplicaciones o recursos de utilidad para la investigación (C4 Gestión técnica).

##### C1. *Referencia especializada*

7. Referencia especializada. Consultas concretas sobre cualquier aspecto relacionado con la investigación. Ejemplo: Pregunta sobre cómo corregir datos en el portal de investigación institucional.
8. Información bibliográfica. Elaboración de bibliografías y selecciones de recursos. Ejemplo: Búsqueda bibliográfica solicitada sobre un tema de investigación.
9. Especialización temática. Consultas especializadas sobre recursos de disciplinas científicas. Ejemplo: Búsqueda solicitada sobre bases de datos específicas de una disciplina.
10. Suministro de documentos. Localización de documentos no disponibles en la universidad. Ejemplo: Entrega de publicaciones solicitadas por personal investigador.
11. Alertas informativas. Demandas individualizadas sobre novedades bibliográficas y de actualidad informativa. Ejemplo: Creación de una alerta informativa personalizada bajo demanda.

##### C2. *Informes*

12. Informes de similitud. Detección de plagios y grado de coincidencia entre documentos. Ejemplo: Comprobación de coincidencias entre un texto nuevo y la bibliografía existente.
13. Selección de revistas. Identificación de publicaciones periódicas para la publicación de resultados de investigación. Ejemplo: Estudio de las principales revistas de un área.
14. Calidad de revistas. Fiabilidad de editoriales académicas y revistas científicas. Ejemplo: Informe sobre la calidad y la fiabilidad de una editorial o revista científica.

15. Licencias de uso. Licencias de comunicación pública de revistas y editoriales científicas. Ejemplo: Datos sobre la licencia de publicación en abierto de una revista.
16. Informes de acreditación. Indicios de calidad de la producción bibliográfica para acreditaciones. Ejemplo: Impacto de publicaciones para un sexenio de investigación.
17. Estudios de productividad. Producción científica de personas o grupos de investigación. Ejemplo: Estudio de las publicaciones de un departamento universitario.
18. Estudios institucionales. Análisis individual o comparativo de la producción científica de la institución. Ejemplo: Informe comparativo de una universidad con otras de sus características.

### *C3. Equipo de trabajo*

19. Revisiones sistemáticas. Revisión bibliográfica de investigaciones. Ejemplo: Elaboración de la revisión bibliográfica para un artículo.
20. Gestión de proyectos. Responsabilidad en proyectos de investigación. Ejemplo: Administración de aplicaciones tecnológicas colaborativas.
21. Archivo en repositorios. Depósito de documentos en repositorios. Ejemplo: Descripción y almacenamiento de resultados de la investigación en repositorios.
22. Gestión de datos. Participación en el plan de datos de investigación. Ejemplo: Carga de datos de investigación en repositorios específicos siguiendo un plan establecido.
23. Gestión editorial. Implicación en los aspectos documentales de los procesos editoriales. Ejemplo: Tratamiento documental de las publicaciones de una universidad.
24. Identificadores de autor. Creación de identificadores persistentes de autor. Ejemplo: Creación de los identificadores de los miembros de un grupo de investigación.
25. Gestión curricular. Introducción de datos en el sistema de gestión curricular de la institución. Ejemplo: Carga de datos curriculares de investigadores en el sistema de gestión de la universidad.
26. Análisis documental. Normalización de resúmenes y lenguajes documentales en artículos. Ejemplo: Redacción de resúmenes y descriptores normalizados en un artículo.
27. Referencias bibliográficas. Elaboración de referencias y citas bibliográficas. Ejemplo: Redacción y revisión de las referencias bibliográficas de una publicación.
28. Divulgación científica. Creación de contenidos divulgativos sobre investigaciones. Ejemplo: Redacción de noticias de divulgación sobre las publicaciones institucionales.

### *C4. Gestión técnica*

29. Portal de la investigación. Gestión de datos en el portal de producción científica. Ejemplo: Administración del portal institucional de producción científica.
30. Repositorio institucional. Administración de comunidades en el repositorio de la universidad. Ejemplo: Administración del repositorio institucional.

31. Plataformas externas. Supervisión de los datos de la producción institucional en servicios externos. Ejemplo: Revisión de datos de investigadores de una universidad en portales de terceros.

#### D. SERVICIOS DE ASISTENCIA

Los usuarios solicitan a la biblioteca orientación sobre recursos, procesos o elementos que intervienen en la investigación. El personal bibliotecario actúa como consultor y asistente en los procesos de la investigación y en los servicios y las herramientas que intervienen en el ciclo de comunicación científica. El apoyo al personal investigador se puede prestar mediante el asesoramiento personalizado (D1 Atención a la investigación) y la ayuda en el uso de espacios, equipos y programas (D2 Soporte informativo).

##### D1. *Atención a la investigación*

32. Atención individual. Asesoramiento personalizado de cualquier aspecto de la comunicación científica. Ejemplo: Recomendación de plataformas para el depósito de datos de investigación.
33. Referencia documental. Soporte en consultas sobre técnicas y recursos documentales. Ejemplo: Asesoramiento sobre datos de la firma institucional.
34. Servicios de autor. Trámites entre personal investigador y las editoriales en las que se quiere publicar. Ejemplo: Intermediación en acuerdos transformativos de publicación.
35. Reputación digital. Apoyo en perfiles de autor y presencia en redes científicas. Ejemplo: Ayuda en la gestión de perfiles de autor en plataformas externas.
36. Consultoría. Colaboración en proyectos y grupos de investigación. Ejemplo: Apoyo en gestión documental en proyectos de investigación.
37. Promoción. Apoyo en la redifusión de contenidos que contribuyan a la expansión de publicaciones institucionales. Ejemplo: Colaboración en los medios de divulgación de la producción de un grupo de investigación.
38. Comunicación institucional. Colaboración con servicios de comunicación institucionales. Ejemplo: Colaboración con los medios de comunicación de la universidad para promocionar resultados de investigación.

##### D2. *Soporte informativo*

39. Referencia técnica. Consultas sobre el uso de programas, tecnologías y plataformas empleadas en la investigación. Ejemplo: Consulta telefónica sobre las licencias de programas informáticos de investigación.
40. Gestión de infraestructuras. Administración de espacios, equipos e instrumentos de investigación. Ejemplo: Gestión de salas de investigación y de préstamo de dispositivos.

## 4. CONCLUSIÓN

Este capítulo identifica y sistematiza los diferentes ámbitos de soporte a la investigación desde las bibliotecas y los inscribe en los diferentes procesos que se desarrollan en la comunicación científica, que se estructura en el modelo DIPE. Los tipos de servicios que las bibliotecas ofrecen para apoyar a sus universidades y al personal investigador en todos los procesos del ciclo de la comunicación científica pueden clasificarse en cuatro grandes modalidades: formación, información, prestación y asistencia. Los servicios de formación incluyen: alfabetización informacional en los procesos y los recursos de la investigación; cursos, tutoriales y actividades instructivas para la adquisición de competencias en comunicación científica. Los servicios de información comprenden: producción de recursos informativos; boletines, guías temáticas, productos informativos, medios y servicios de comunicación que informen sobre recursos documentales, técnicas de investigación, publicación de resultados y difusión de la producción científica. Los servicios de prestación contienen: elaboración de productos bibliográficos y documentales para la investigación; búsquedas, informes y estudios, información elaborada y servicios requeridos en los procesos de la comunicación científica y la evaluación de la ciencia. Por último, los servicios de asistencia engloban: colaboración en el desarrollo de la investigación; servicios de consultoría, apoyo funcional a la gestión de la investigación y la administración de plataformas; asistencia en relación con infraestructuras, equipamientos, aplicaciones, normativas y políticas de investigación. Los servicios de apoyo a la comunicación científica se pueden prestar de diferentes formas. Este capítulo identifica cuarenta tipos de servicios, organizados en las cuatro modalidades, que pueden servir como una guía de posibilidades que las bibliotecas universitarias pueden tener presentes en la planificación de sus servicios especializados de apoyo a los procesos de la comunicación científica.

## REFERENCIAS BIBLIOGRÁFICAS

- Atkinson, J. (2016). Academic Libraries and Research Support: An Overview. En J. Atkinson (Ed.), *Quality and the Academic Library* (pp. 135-141). Chandos.
- Brown, M. L. (2013). The role of the research library. En D. Shorley y M. Jubb (Eds.), *The Future of Scholarly Communication* (pp. 157-168). Facet. <https://doi.org/10.29085/9781856049610.014>
- Chigwada, J. P. (2020). The Role of the Librarian in the Research Life Cycle: Research Collaboration Among the Library and Faculty. En C. T. Chisita (Ed.), *Advances in Library and Information Science* (pp. 335-346). IGI Global. <https://doi.org/10.4018/978-1-7998-0043-9.ch017>
- Comisión Asesora de Bibliotecas y Documentación. (2020). *Servicios de apoyo a la investigación y publicación científica*. CABID CRUCH. <http://cabid.cl/wp-content/uploads/2023/07/Servicios-de-apoyo-a-la-investigacion-y-publicacion-cientifica.pdf>
- Edinburgh Napier University. (2024). *Research Cycle*. <https://staff.napier.ac.uk/services/information-services/research-cycle/Pages/home.aspx>
- Fernández-Marcial, V. y González-Solar, L. (Eds.). (2021). *Cases on Research Support Services in Academic Libraries*. IGI Global. <https://doi.org/10.4018/978-1-7998-4546-1>

- Hoffman, S. (2016). *Dynamic Research Support for Academic Libraries*. Facet. <https://doi.org/10.29085/9781783301225>
- Jiang, J., Maurici-Pollock, D. y Tang, R. (2025). Research Data Lifecycle (RDLC): An Investigation into the Disciplinary Focus, Use Cases, Creator Backgrounds, Stages and Shapes of RDLC Models. *International Journal of Digital Curation*, 19(1), 30. <https://doi.org/10.2218/ijdc.v19i1.860>
- Jubb, M. (2016). Libraries and the Support of University Research. En J. Atkinson (Ed.), *Quality and the Academic Library* (pp. 143-156). Chandos. <https://doi.org/10.1016/B978-0-12-802105-7.00014-2>
- Maron, N., Kennison, R., Bracke, P., Hall, N., Gilman, I., Malenfant, K., Roh, C. y Shorish, Y. (2019). *Open and Equitable Scholarly Communications: Creating a More Inclusive Future*. Association of College and Research Libraries. <https://doi.org/10.5860/acrl.1>
- Merlo Vega, J. A. (2022). Los servicios especializados en comunicación científica: Una urgente prioridad de la biblioteca universitaria. En *Reunión de Bibliotecarios de la Península de Yucatán (28. 2022. Mérida)*. XXVIII Reunión de Bibliotecarios de la Península de Yucatán, Mérida (México). <http://hdl.handle.net/10366/149113>
- Merlo Vega, J. A. (2024). Servicios bibliotecarios especializados en investigación: Competencias, modalidades y catálogo de servicios. *Revista Española de Documentación Científica*, 47(4), e401. <https://doi.org/10.3989/redc.2024.4.1639>
- Ojanen, M., Lindholm, T. y Siipilehto, L. (2020, septiembre 3). What is research data management (RDM)? *Think Open*. <https://blogs.helsinki.fi/thinkopen/know-your-data-rdm-series-1/>
- Ragon, B. (2019). Alignment of library services with the research lifecycle. *Journal of the Medical Library Association*, 107(3). <https://doi.org/10.5195/jmla.2019.595>
- REBIUN. (2024). *V Plan estratégico REBIUN: 2024-2027*. Red de Bibliotecas Universitarias y Científicas Españolas. <https://hdl.handle.net/20.500.11967/1345>
- Rey Martín, C., Camón Luis, E. y Pacheco, F. (2018). El soporte a la investigación en las bibliotecas universitarias españolas. *Anales de Documentación*, 21(1). <http://dx.doi.org/110.6018/analesdoc.21.1.295841>
- San-José-Montano, B., Fernández-Fernández-Cuesta, P. y Espantaleón-Agreda, M. (2025). Papel de los bibliotecarios especializados en la comunicación científica: Formación en buenas prácticas y en el uso ético de la información. En N. R. Dinu y T. Baiget (Eds.), *Divulgación, transferencia e impacto social de la ciencia* (pp. 130-160). Ediciones Profesionales de la Información. <https://doi.org/10.3145/codi2025/001>
- Tran, N.-Y. y Chan, E. K. (2020). *Supporting Scholarly Research: Current and New Opportunities for Academic Libraries*. Association of College & Research Libraries. <http://choice360.org/librarianship/whitepaper>
- University of Central Florida Libraries. (2021). *Research Lifecycle. Scholarly Communication*. <https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/>
- Werner, I. M. (2017). *From access to accessibility: The university library of the future in the scholarly communication cycle*. Utrecht University Library. <https://doi.org/10.33540/UU1>



## LOS DATASETS EN ABIERTO Y SU APORTACIÓN AL IMPACTO SOCIAL DE LA CIENCIA

### OPEN DATASETS AND THEIR CONTRIBUTION TO THE SOCIAL IMPACT OF SCIENCE

Paloma G. BLÁZQUEZ<sup>1</sup>, Almudena MANGAS-VEGA<sup>2</sup> y Francisco José GARCÍA-PEÑALVO<sup>3</sup>

<sup>1</sup> *Universidad de Salamanca. España*

palomagb@usal.es

 <https://orcid.org/0009-0002-6986-0490>

<sup>2</sup> *Universidad de Salamanca. España*

almumvega@usal.es

 <https://orcid.org/0000-0002-3464-3624>

<sup>3</sup> *Universidad de Salamanca. España*

fgarcia@usal.es

 <https://orcid.org/0000-0001-9987-5584>

RESUMEN: Los *datasets* en abierto se han consolidado como una herramienta clave dentro de la ciencia abierta por su potencial para generar impacto más allá del ámbito académico; y, a su vez, como aportación y/o publicación científica susceptible de ser evaluada en convocatorias regionales y nacionales. Este trabajo presenta un estudio de caso centrado en los repositorios institucionales de las universidades públicas de Castilla y León, con el objetivo de analizar el alcance de estos conjuntos de datos no solo como análisis, sino también en términos de visibilidad e impacto social. A partir de métricas de uso (visitas, descargas) y altmétricas disponibles (PlumX, Dimensions), se observa una presencia incipiente pero significativa en términos de reutilización social y académica. Se analiza también la alineación con los principios FAIR, así como las posibilidades que estos *datasets* ofrecen para la innovación ciudadana, la mejora de políticas públicas o la educación informal. Finalmente, se discuten posibles líneas de acción para favorecer la comunicación, la visibilidad y la reutilización de los datos de investigación en abierto como vía para potenciar su impacto social.

PALABRAS CLAVE: *datasets* abiertos; impacto social; ciencia abierta; repositorios institucionales; métricas de uso; principios FAIR.

**ABSTRACT:** Open datasets have become a key tool within Open Science due to their potential to generate impact beyond academia. This study presents a case study focused on the institutional repositories of public universities in Castilla y León (Spain), aiming to analyse the social impact of these datasets. Based on usage metrics (views, downloads) and available altmetrics (PlumX, Dimensions), the results show an emerging but significant presence in terms of academic and social reuse. The analysis also includes the alignment with FAIR principles and the opportunities these datasets offer for citizen innovation, policy development and informal education. Lastly, future directions are discussed to enhance communication, visibility and reuse of open research data as a way to foster their societal impact.

**KEYWORDS:** open datasets; social impact; open science; institutional repositories; usage metrics; FAIR principles.

## 1. INTRODUCCIÓN

El impacto de la ciencia se ha consolidado como uno de los principales retos para los sistemas de investigación –y evaluación de la investigación– contemporáneos. La creciente exigencia de que la ciencia contribuya de forma tangible al bienestar social ha situado en el centro del debate no solo qué se investiga, sino cómo se comunica y se hace accesible ese conocimiento. Ya no basta con generar resultados: es imprescindible que sean comprensibles, útiles y reutilizables por diferentes actores, dentro y fuera del ámbito académico.

En este contexto, la ciencia abierta representa un cambio de paradigma que transforma no solo la producción, sino también la difusión y la evaluación del conocimiento científico. Este enfoque trasciende la mera apertura de datos o publicaciones, promoviendo una práctica investigadora basada en la transparencia, la accesibilidad y la colaboración, y cuestionando modelos tradicionales más cerrados y jerárquicos.

Como subrayan Burgelman *et al.* (2019), avanzar hacia una ciencia más abierta exige también adoptar prácticas sistemáticas de documentación, trazabilidad y preservación de los datos, incluso cuando no se contemple su apertura total. En este marco, los principios FAIR (Findable, Accessible, Interoperable, Reusable) sintetizan la aspiración de construir un ecosistema científico más eficiente, responsable y alineado con las necesidades sociales.

Uno de los pilares de esta transformación hacia una ciencia más abierta lo constituyen los repositorios institucionales, que han dejado de ser meras plataformas de almacenamiento para consolidarse como infraestructuras estratégicas en la visibilización, la circulación y la evaluación de los resultados científicos. Al integrar métricas de uso –como visualizaciones o descargas– junto con herramientas alométricas, permiten rastrear nuevas formas de impacto que trascienden el ámbito estrictamente académico (Khan *et al.*, 2021).

En este contexto, los datos abiertos dejan de considerarse un subproducto del proceso investigador para posicionarse como un recurso clave con potencial transformador en múltiples niveles (Fecher y Friesike, 2013). Facilitan la transferencia de conocimiento y su aplicación en distintos sectores, y resultan esenciales para promover la toma de decisiones

basada en la evidencia, el diseño de políticas públicas, el impulso de la innovación –social y tecnológica– y el fortalecimiento de la educación científica. Así, los repositorios institucionales no solo refuerzan la transparencia y la accesibilidad de la ciencia, sino que también amplifican su impacto social.

Esta concepción ha ido integrándose en las políticas científicas a través de marcos normativos que reconocen el valor de los datos como resultados científicos en sí mismos. En España, este reconocimiento se refleja en convocatorias como el programa *Generación del Conocimiento* financiado por la Agencia Estatal de Investigación (AEI)<sup>1</sup>, donde los Planes de Gestión de Datos (PGD) son ya un requisito obligatorio. A nivel europeo, programas como Horizon Europe<sup>2</sup> refuerzan esta perspectiva, estableciendo el acceso abierto a los datos como un principio estructural de la investigación financiada con fondos públicos. La Comisión Nacional Evaluadora de la Actividad Investigadora (CNEAI)<sup>3</sup> exige que los conjuntos de datos estén depositados en repositorios de acceso abierto, de acuerdo con lo establecido en el artículo 37.2 de la Ley 17/2022 de la Ciencia, la Tecnología y la Innovación, para que puedan ser considerados en los procesos de evaluación.

Esta evolución normativa no solo afecta a las obligaciones de los proyectos financiados, sino que también refleja un cambio más profundo en los criterios de evaluación científica. Compromisos internacionales como la Declaración de San Francisco sobre la Evaluación de la Investigación (DORA) o el Acuerdo sobre la Reforma de la Evaluación de la Investigación (COARA) cuestionan la dependencia de indicadores tradicionales centrados en el prestigio de las publicaciones y apuestan por sistemas más justos y centrados en el valor real de los resultados. En línea con estos enfoques, los conjuntos de datos empiezan a ser reconocidos como resultados científicos legítimos, especialmente por su contribución a una ciencia más abierta, reutilizable y conectada con su entorno. Esta ampliación del espectro de resultados evaluables permite considerar impactos que trascienden lo académico. Como señalan Penfield *et al.* (2014), la investigación puede generar efectos instrumentales, sociales, culturales y políticos que repercuten directamente en la economía, las políticas públicas, la salud o la calidad de vida.

No obstante, el reconocimiento formal de estos nuevos tipos de resultados científicos, como los conjuntos de datos, no implica por sí solo su efectiva incorporación en los marcos de evaluación ni garantiza su aprovechamiento real en los procesos de toma de decisiones,

1 Agencia Estatal de Investigación. (2024). *Convocatoria de ayudas a «Proyectos de Generación de Conocimiento» y actuaciones para la formación de personal investigador predoctoral asociadas a dichos proyectos*. [https://www.aei.gob.es/sites/default/files/convocatory\\_info/file/2024-12/Convocatoria\\_ProyectosGeneraci%C3%B3nConocimiento\\_PID2024\\_%20%281%29.pdf](https://www.aei.gob.es/sites/default/files/convocatory_info/file/2024-12/Convocatoria_ProyectosGeneraci%C3%B3nConocimiento_PID2024_%20%281%29.pdf)

2 OpenAIRE. (s. f.). *How to comply with Horizon Europe mandate for Research Data Management*. <https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm>

3 Ministerio de Ciencia, Innovación y Universidades. (2024, 24 de diciembre). Resolución de 11 de diciembre de 2024, de la Secretaría General de Universidades, por la que se aprueba la convocatoria de evaluación de la actividad investigadora. *BOE* núm. 309, pp. 179840-179845. Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA).

[https://www.aneca.es/documents/20123/238516/BOE-A-2024-27007\\_convotaria+SGU.pdf/943922c6-1f98-6ac4-063c-d1b1788324eb?t=1736258650652](https://www.aneca.es/documents/20123/238516/BOE-A-2024-27007_convotaria+SGU.pdf/943922c6-1f98-6ac4-063c-d1b1788324eb?t=1736258650652)

transferencia o reutilización. A pesar de estos avances normativos y conceptuales, persisten importantes desafíos a la hora de evaluar de forma efectiva el impacto de los datos abiertos. Que existan políticas que promuevan su disponibilidad no implica necesariamente que se esté logrando una circulación real ni un aprovechamiento significativo de estos recursos. La trazabilidad del impacto sigue siendo limitada, en parte por la falta de indicadores específicos que permitan capturar su reutilización, su alcance más allá del ámbito académico o su integración en otros contextos sociales, económicos o institucionales. Como resultado, sigue siendo difícil determinar hasta qué punto los *datasets* están contribuyendo de forma tangible a la transferencia de conocimiento o a la toma de decisiones basada en evidencia.

Este trabajo parte del análisis del uso de los repositorios institucionales de las universidades públicas de Castilla y León, con el objetivo de identificar patrones, tendencias y diferencias en cuanto al volumen de depósitos, la frecuencia de uso y el grado de adopción del acceso abierto en cada institución. A través de esta aproximación, se busca comprender en qué medida los repositorios forman parte de la práctica investigadora cotidiana y si existe una cultura consolidada —o aún emergente— en torno al depósito de datos científicos.

A partir de este primer análisis, el estudio se centra en los conjuntos de datos depositados, combinando métricas de uso (como visualizaciones y descargas) con indicadores alométricos para caracterizar su visibilidad digital y su circulación más allá del ámbito académico. Este enfoque permite explorar no solo su nivel de reutilización y presencia en otros contextos, sino también si estos datos están generando un impacto real o aún incipiente en términos de transferencia y comunicación científica.

## 2. METODOLOGÍA

La presente investigación adopta una aproximación metodológica de carácter cuantitativo, orientada al análisis de la visibilidad, el uso y el impacto potencial de los conjuntos de datos depositados en acceso abierto a través de repositorios institucionales universitarios. El estudio se articula dentro de una estrategia no experimental, de tipo exploratorio-descriptivo, cuyo objetivo principal es identificar patrones de circulación y reutilización de los *datasets*, así como evaluar su presencia en espacios alternativos de comunicación científica y su proyección en términos de impacto académico y social.

La unidad de análisis está constituida por los conjuntos de datos depositados en los repositorios institucionales de las universidades públicas que integran la red BUCLE (Bibliotecas Universitarias de Castilla y León)<sup>4</sup>. La selección de los registros se realizó de forma exhaustiva, incluyendo todos los documentos identificados explícitamente como conjuntos de datos depositados en los repositorios institucionales de las universidades públicas que integran el Consorcio BUCLE (Bibliotecas Universitarias de Castilla y León).

<sup>4</sup> Consorcio de Bibliotecas Universitarias de Castilla y León (BUCLE). (s. f.). *Consorcio BUCLE*. <https://consorcio-bucle.es/>

De forma paralela también se analizan, por observación, otros elementos estructurales como la normalización de campos, metadatos, funcionalidades y datos ofrecidos, así como las diferencias entre los diferentes repositorios de la red BUCLE.

La recopilación de datos incluyó variables agrupadas en tres dimensiones analíticas:

- Dimensión descriptiva: recoge información básica sobre cada conjunto de datos, incluyendo título, autores, institución de filiación, fecha de publicación, área de conocimiento y tipo de documento.
- Dimensión de uso: analiza las métricas de interacción disponibles en los propios repositorios institucionales, concretamente el número total de visualizaciones y descargas. Estos indicadores se interpretan como una primera aproximación al interés que generan los *datasets* en entornos digitales.
- Dimensión altmétrica: contempla señales de atención provenientes de fuentes no tradicionales, recopiladas a través de las plataformas PlumX y Dimensions. Se incluyen menciones en redes sociales (Twitter/X, Facebook), gestores de referencia (Mendeley, Zotero), blogs académicos, noticias en medios de comunicación, documentos de política pública y otras formas de circulación que trascienden el ámbito estrictamente bibliométrico.

La extracción de datos se realizó entre el 1 y el 20 de abril de 2025, accediendo directamente a las páginas web de los repositorios institucionales de cada universidad pública de Castilla y León. El proceso combinó revisión manual y recolección semiautomatizada, con el fin de garantizar la homogeneidad de las variables consideradas y la calidad de los registros incluidos.

Los datos fueron organizados en una base estructurada y analizados mediante técnicas estadísticas descriptivas (frecuencias, medias, desviaciones estándar), así como análisis exploratorios de correlación. El objetivo fue detectar tendencias, observar los niveles de atención que reciben los *datasets* y explorar posibles relaciones entre visibilidad y reutilización.

De forma complementaria, se analiza la relación entre estos indicadores y métricas bibliométricas tradicionales, con el fin de valorar en qué medida la visibilidad en acceso abierto de los conjuntos de datos se asocia con un mayor impacto en términos de citación y reconocimiento académico de la producción científica relacionada, tal como sugieren estudios previos (García-Blázquez, 2024; Peter *et al.*, 2016).

Como parte del análisis contextual, se consideraron también aspectos cualitativos referidos al grado de cumplimiento de los principios FAIR, la normalización de los metadatos y la interoperabilidad de los sistemas de información. Estos elementos se contemplan como variables de entorno que pueden condicionar la trazabilidad, la visibilidad y la potencial reutilización de los conjuntos de datos analizados.

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. DIMENSIÓN DESCRIPTIVA: CARACTERIZACIÓN DE LOS REPOSITORIOS INSTITUCIONALES

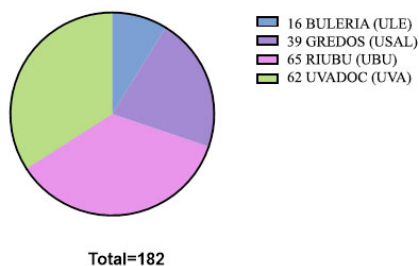
El análisis se ha centrado en los repositorios institucionales de las universidades públicas que integran el consorcio BUCLE (Bibliotecas Universitarias de Castilla y León), formado por la Universidad de Burgos (UBU), la Universidad de León (ULE), la Universidad de Salamanca (USAL) y la Universidad de Valladolid (UVA). Estas instituciones comparten un marco normativo regional y una estructura de gobernanza académica similar, pero presentan diferencias relevantes en términos de cultura institucional en torno a la ciencia abierta, grado de desarrollo de los servicios de apoyo a la gestión de datos y volumen general de producción científica.

En total, se identificaron 182 conjuntos de datos publicados en los respectivos repositorios institucionales: RIUBU (UBU), UVaDOC (UVA), GREDOS (USAL) y BULERIA (ULE). Como muestra la Figura 1, la distribución de los *datasets* es desigual. El mayor número de registros se concentra en RIUBU, el repositorio de la Universidad de Burgos ( $n = 65$ ), seguido muy de cerca por UVaDOC ( $n = 62$ ), el repositorio de la Universidad de Valladolid. En cambio, GREDOS, de la Universidad de Salamanca, contiene 39 conjuntos de datos, mientras que BULERIA, de la Universidad de León, alberga solo 16. Esta heterogeneidad inicial pone de manifiesto no solo la diferente adopción institucional de prácticas de apertura de datos, sino también las prioridades de desarrollo técnico y documental que cada universidad ha establecido en relación con sus repositorios.

Los repositorios analizados permiten el depósito de conjuntos de datos de forma explícita, ya sea mediante tipologías documentales específicas o a través de etiquetas aplicadas en procesos de autoarchivo. No obstante, a pesar de estar integrados en un mismo consorcio (BUCLE), la información disponible sobre los *datasets* no es homogénea entre instituciones. Existen diferencias sustanciales en cuanto al tipo de información que se recoge, los criterios de clasificación, la visibilidad del recurso dentro del repositorio y el nivel de estandarización de los registros. En algunos casos, los conjuntos de datos aparecen integrados en colecciones generales sin distinción clara respecto a otros documentos, o bien se encuentran etiquetados bajo denominaciones genéricas que dificultan su identificación y recuperación. Estas divergencias afectan directamente a la trazabilidad de los recursos y dificultan tanto su localización como su posterior reutilización.

Los datos recogidos en este estudio reflejan algunas limitaciones estructurales en la gestión de conjuntos de datos. Aunque el 96,65 % de los registros analizados cuenta con un identificador persistente (DOI), un 3,35 % sigue utilizando el sistema HANDLE, menos interoperable en ciertos contextos científicos y métricos. Esta falta de uniformidad en la identificación puede afectar a la fiabilidad del recurso y dificultar su integración en sistemas de evaluación y seguimiento del impacto.

Figura 1



Fuente: Elaboración propia a partir de datos de los repositorios de BUCLE (n = 182).

La situación es aún más llamativa en relación con los *datapapers*: el 32,60 % de los conjuntos de datos no está acompañado de ningún documento que los describa o contextualice. Lejos de ser un elemento accesorio, el *datapaper* es clave para garantizar que los datos sean comprensibles, trazables y reutilizables, especialmente en contextos interdisciplinares o no especializados.

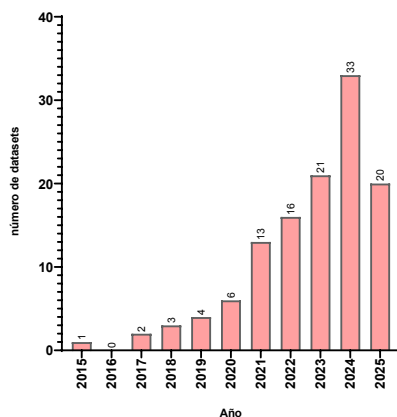
Estas carencias apuntan a una implementación todavía parcial de los principios FAIR en los repositorios institucionales. Aunque algunos requisitos formales, como el uso de DOI, se cumplen mayoritariamente, no siempre van acompañados de una documentación adecuada que permita aprovechar los datos de forma efectiva. Como destacan Cousijn *et al.* (2018), la combinación de identificadores persistentes con una buena descripción no solo mejora la visibilidad, sino que es fundamental para garantizar la transparencia y la reutilización. Fortalecer estos aspectos es clave para avanzar hacia una cultura de datos abiertos que vaya más allá del cumplimiento normativo y se centre en la utilidad real de los datos para la comunidad científica y la sociedad.

La evolución temporal del número de conjuntos de datos depositados en acceso abierto entre 2015 y 2025 muestra una tendencia sostenidamente creciente, reflejo de una adopción progresiva –aunque aún no plenamente consolidada– de prácticas vinculadas a la ciencia abierta en el contexto universitario regional. Entre 2015 y 2019, el número de *datasets* depositados en los repositorios institucionales de las universidades públicas de Castilla y León fue testimonial ( $n \leq 4$ ), lo que sugiere una fase inicial de baja implantación, posiblemente relacionada con la falta de políticas específicas, estructuras de soporte técnico o incentivos claros para el personal investigador.

A partir de 2020 se detecta una inflexión en la curva, con un crecimiento sostenido que alcanza un primer punto de aceleración en 2021 ( $n = 13$ ) y que culmina en 2024 con un máximo histórico de 33 conjuntos de datos publicados. Si bien los datos correspondientes a 2025 se encuentran aún en fase de recopilación, ya se contabilizan 20 *datasets* depositados hasta abril, lo que permite anticipar, de mantenerse la tendencia, una cifra final superior a la del año anterior. Esta progresión, más allá del aumento cuantitativo, puede interpretarse como

un indicio de consolidación de los repositorios institucionales como infraestructuras válidas y reconocidas para la publicación de resultados de investigación en forma de datos abiertos.

Figura 2



Fuente: Elaboración propia a partir de datos de BUCLE (2015-2025, n = 182).

Este patrón no responde únicamente a decisiones individuales, sino que se enmarca en un contexto institucional que favorece la publicación de datos como resultado científico evaluable. El creciente protagonismo del *datapaper* refleja esta tendencia, consolidándose como una herramienta clave para reforzar la trazabilidad, la visibilidad y la reutilización de los conjuntos de datos (Cole *et al.*, 2024). La progresiva implementación de marcos como el artículo 37.2 de la Ley de la Ciencia, así como la adopción de los principios FAIR, han contribuido a generar un entorno más receptivo a la gestión responsable de los datos, incentivando nuevas formas de legitimación académica vinculadas a su publicación y circulación.

El análisis de las autorías asociadas a los conjuntos de datos revela una fuerte concentración institucional y una baja recurrencia en la publicación de datos abiertos. Del total de 470 autorías identificadas, la mayoría se concentra en la Universidad de Burgos (n = 207), seguida por la Universidad de Valladolid (n = 130), la Universidad de Salamanca (n = 98) y, en menor medida, la Universidad de León (n = 35). Esta distribución sugiere diferencias significativas entre instituciones, posiblemente vinculadas al apoyo institucional, la existencia de equipos especializados o el grado de integración de la ciencia abierta en las dinámicas de investigación.

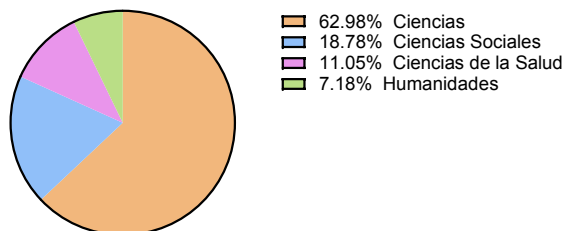
En cuanto a la frecuencia de contribución, el patrón es claramente puntual: el 75,3 % de los autores ha participado en un único depósito. Solo una minoría firma más de dos conjuntos de datos y los casos con diez o más aportaciones son excepcionales. Este comportamiento sugiere que el depósito de datos todavía no forma parte habitual de la trayectoria investigadora. La media de autoría por *dataset* es de 4,2 personas (mediana = 4; moda = 2), lo que indica una colaboración moderada. Sin embargo, existe una gran variabilidad: la mayoría de los

conjuntos tienen pocos autores, pero hay algunos casos excepcionales con una participación mucho más amplia. La distribución de autorías está claramente sesgada hacia los valores bajos, con una concentración elevada en un número reducido de contribuciones. Lo observado pone de manifiesto que el depósito de datos sigue siendo una práctica poco extendida y repartida de forma desigual entre los investigadores.

Por otro lado, el análisis de las áreas de conocimiento permite identificar en qué disciplinas el depósito de datos en acceso abierto está más consolidado y dónde sigue siendo una práctica minoritaria. Esto resulta clave para detectar patrones de adopción y posibles desigualdades entre campos científicos. Diversos estudios han señalado que la reproducibilidad científica depende en gran medida de la disponibilidad de datos abiertos, especialmente en disciplinas como la medicina, la genética evolutiva o las ciencias biológicas, donde estas prácticas están más asentadas (Anagnostou *et al.*, 2013). A partir de esta premisa, se examinó la distribución de los conjuntos de datos por macroáreas y áreas de conocimiento.

Los resultados muestran que más del 60 % de los *datasets* analizados pertenecen al ámbito de las ciencias experimentales y naturales, seguidas, a notable distancia, por las ciencias sociales, las ciencias de la salud y las humanidades. Esta distribución no solo refleja los volúmenes diferenciales de producción científica entre campos, sino también las asimetrías estructurales en la adopción de prácticas de ciencia abierta, particularmente en lo que respecta al manejo, la documentación y la publicación de datos primarios.

Figura 3



Fuente: Elaboración propia según área de conocimiento (n = 182).

A nivel de áreas específicas, el depósito de datos se concentra en disciplinas con tradición cuantitativa y requerimientos claros de apertura por parte de agencias financiadoras, como la química, la educación o la informática. Este patrón muestra que el avance hacia una cultura de datos abiertos sigue lógicas disciplinares, influido por factores como la disponibilidad de infraestructuras, normativas específicas o el grado de familiaridad con los principios FAIR. Esta concentración condiciona la evaluación del impacto: al centrarse en campos concretos, los análisis de visibilidad y reutilización no reflejan la diversidad del sistema científico. Por ello, es fundamental promover la apertura en áreas menos representadas y desarrollar métricas sensibles al contexto, que reconozcan las distintas formas en que cada disciplina contribuye al conocimiento y a la sociedad.

### 3.2. DIMENSIÓN DE USO: PRIMEROS INDICIOS DE CIRCULACIÓN E INTERÉS SOCIAL

Tras caracterizar los repositorios desde el punto de vista institucional y documental, resulta clave analizar en qué medida los conjuntos de datos están siendo realmente consultados por la comunidad científica u otros usuarios potenciales. Para ello, se ha incorporado un análisis de la dimensión de uso, centrado en dos indicadores básicos: el número de visualizaciones y descargas registrado por los propios repositorios. Aunque estas métricas no permiten medir directamente el impacto social, sí ofrecen una primera señal de interés y circulación que ayuda a observar patrones de visibilidad y alcance.

Desde una perspectiva metodológica, estas cifras se interpretan como indicadores de atención: pequeñas huellas cuantificables que reflejan interacciones reales con los datos. Su valor reside precisamente en esa función basal. La ausencia total de visualizaciones o descargas puede indicar una nula proyección digital, mientras que su presencia –aunque moderada– sugiere un posible uso, consulta o reutilización. Este tipo de interacción puede entenderse como una forma incipiente de impacto, aún lejos de transformaciones consolidadas, pero necesaria para que estas lleguen a producirse.

Algunos repositorios ofrecen, además, datos desagregados sobre el origen geográfico de los accesos –por países, ciudades o instituciones–, así como mapas interactivos con visualizaciones geolocalizadas. Esta información aporta una capa interpretativa adicional, al permitir analizar no solo la intensidad, sino también la distribución territorial del interés generado. Aunque estas métricas tienen limitaciones (como el posible efecto del tráfico automatizado o de *bots*), su análisis crítico aporta una aproximación válida para explorar la proyección externa de los datos. En definitiva, aunque todavía desigual, la dimensión de uso representa un primer paso para pensar el impacto social de los datos abiertos y refuerza la importancia de integrar este tipo de señales en los marcos de evaluación científica y de transferencia del conocimiento.

En términos cuantitativos, las métricas de uso revelan una circulación desigual: aunque el promedio por conjunto es de aproximadamente 268 visualizaciones y 142 descargas, estos valores esconden una alta disparidad. Las distribuciones son claramente asimétricas y leptocúrticas –con coeficientes superiores a 3 en asimetría y más de 15 en curtosis–, lo que implica que unos pocos *datasets* concentran la mayor parte de la atención, mientras que muchos otros apenas generan interacción. Lejos de tratarse de casos anecdóticos, estos focos de interés intensivo merecen un análisis cualitativo más profundo para identificar qué factores –temática, visibilidad, documentación, contexto institucional– podrían estar detrás de su mayor proyección.

Entre los repositorios analizados, se observan diferencias notables. La Universidad de Valladolid (UVaDOC) destaca por su mayor nivel medio de interacciones, lo que sugiere un entorno más favorable para la circulación activa de los datos. En contraste, repositorios como GREDOS (USAL) o BULERIA (ULE), pese a contar con volúmenes relevantes de depósito, presentan cifras medias más bajas. Esta variabilidad sugiere que la infraestructura y las políticas institucionales pueden influir significativamente en la visibilidad digital de los conjuntos de datos.

Tabla 1

Visitas		Descargas	
Media	268,29	Media	142,41
Error típico	16,41	Error típico	26,44
Mediana	225,00	Mediana	51,00
Moda	149,00	Moda	20,00
Desviación estándar	220,75	Desviación estándar	354,75
Varianza de la muestra	48728,88	Varianza de la muestra	125845,62
Curtosis	15,32	Curtosis	57,15
Coefficiente de asimetría	3,26	Coefficiente de asimetría	6,81
Rango	1750,00	Rango	3603,00
Mínimo	10,00	Mínimo	1,00
Máximo	1760,00	Máximo	3604,00
Suma	48560,00	Suma	25634,00
Cuenta	182	Cuenta	182

Fuente: Elaboración propia a partir de los datos de uso y descarga de *datasets* (n = 181).

Si bien estos resultados no evidencian aún un impacto social consolidado –entendido como influencia directa en políticas públicas, prácticas profesionales o transformaciones sociales–, sí permiten identificar un escenario de activación potencial. En contextos de ciencia abierta, visualizaciones y descargas son trazas valiosas: no prueban una repercusión directa, pero sí reflejan movimiento, interés y posibilidad de uso. Incorporar esta dimensión en los marcos de evaluación resulta esencial para construir indicadores más ajustados al papel real de los datos en los procesos de transferencia y conexión con públicos diversos.

### 3.3. DIMENSIÓN ALTMÉTRICA: COMUNICAR PARA REUTILIZAR Y TRANSFORMAR

Tras haber examinado las métricas de uso vinculadas al acceso a los conjuntos de datos, resulta necesario dar un paso más en la comprensión de su circulación. El acceso, aunque necesario, no constituye por sí solo una manifestación de impacto. En el contexto de la ciencia abierta, y especialmente bajo el marco de los principios FAIR, la mera disponibilidad del dato no garantiza su reutilización ni su incidencia real en la transformación del entorno. Para que los datos abiertos puedan convertirse en vehículos efectivos de impacto social, deben ser no solo localizables y accesibles, sino también comprensibles, trazables y comunicados de forma estratégica. Esto implica no solo una adecuada estructuración documental, sino también una proyección comunicativa que les permita llegar a públicos diversos, con diferentes niveles de especialización y objetivos de uso.

Desde esta perspectiva, pueden distinguirse dos formas principales de reutilización de los datos abiertos. Por un lado, la reutilización académica, que incluye su citación, integración en nuevas investigaciones, uso docente o replicación de estudios. Por otro, la reutilización social, orientada a su aplicación en contextos no académicos, como políticas públicas, innovación ciudadana o divulgación científica. Ambas vías son complementarias y necesarias, pero responden a lógicas y circuitos distintos.

En lo que respecta al ámbito académico, la citación se mantiene como uno de los indicadores más sólidos para rastrear la incorporación de un conjunto de datos en nuevas publicaciones. Los repositorios integrados en el Consorcio BUCLE disponen de mecanismos automatizados para el rastreo de citas, utilizando fuentes como Dimensions para identificar referencias académicas. Sin embargo, los datos analizados muestran que su uso en nuevas investigaciones sigue siendo muy bajo: solo 5 de los 182 conjuntos estudiados tienen citas registradas, y todos ellos pertenecen al repositorio GREDOS de la Universidad de Salamanca. Esta cifra, que representa solo el 2,75 % del total, refleja la escasa reutilización académica efectiva de los datos disponibles.

Este resultado no parece explicarse únicamente por un posible desfase temporal entre la publicación de los conjuntos de datos y su eventual citación. Algunos de ellos llevan más de cinco años disponibles sin haber sido referenciados en publicaciones científicas. Todo apunta, por tanto, a la existencia de barreras estructurales que dificultan una reutilización efectiva: desde la limitada visibilidad de muchos *datasets*, hasta la ausencia de documentación adecuada —como los *datapapers*— o una cultura de citación de datos aún poco consolidada en determinadas disciplinas. La citación no solo permite rastrear el uso académico de los datos, sino que constituye un mecanismo clave de reconocimiento y trazabilidad. Su escasa presencia en este estudio pone de relieve los desafíos que aún persisten para consolidar prácticas sostenidas de reutilización académica en el contexto de la ciencia abierta.

Por otro, la reutilización social, que se refiere a la aplicación de los datos en políticas públicas, procesos de innovación ciudadana, divulgación en medios o actividades de alfabetización científica. Ambas vías resultan necesarias y complementarias, aunque se desarrollan en contextos distintos y responden a lenguajes, tiempos y destinatarios específicos.

Esta ampliación del enfoque sobre el valor de los datos se inscribe en una transformación más amplia de los sistemas internacionales de evaluación científica. Iniciativas como DORA y CoARA impulsan marcos más inclusivos, que trascienden las métricas centradas exclusivamente en la producción académica, y reivindican una valoración más rica del impacto, incluyendo su potencial transformador en esferas sociales, económicas, políticas o medioambientales.

En línea con este cambio, han adquirido protagonismo indicadores que permiten captar señales de circulación e influencia en espacios no estrictamente académicos. Entre ellos, destacan las altmétricas, concebidas como herramientas para detectar formas tempranas de atención hacia la ciencia en entornos digitales. Como señala Thelwall (2018), estos indicadores surgieron para registrar interacciones que escapan a los sistemas de citación tradicionales, ofreciendo una imagen más diversa del modo en que la investigación empieza a difundirse y discutirse.

En el caso de los datos abiertos, su diseño con identificadores persistentes, metadatos estandarizados y sistemas interoperables permite rastrear su circulación digital con mayor precisión. Esta trazabilidad técnica habilita el uso de alométricas que capturan indicios de visibilidad en plataformas como redes sociales, blogs, medios digitales o foros científicos colaborativos. Aunque estas señales no constituyen por sí mismas un impacto social consolidado, sí actúan como marcadores tempranos de atención e interés.

Desde esta perspectiva, las alométricas permiten observar si los *datasets* comienzan a insertarse en dinámicas de circulación que podrían derivar en formas de apropiación social. Funcionan, por tanto, como termómetros iniciales: no prueban que se haya producido un cambio, pero sí permiten detectar si los datos están empezando a generar conexiones significativas fuera del entorno académico.

A nivel instrumental, existen diversas plataformas diseñadas para rastrear señales de atención en línea asociadas a los objetos digitales de investigación. Herramientas como Altmetric.com, PlumX Metrics, Dimensions, Overton o Lens.org permiten monitorizar este tipo de interacciones en tiempo real, cada una con distintos enfoques metodológicos, fuentes y criterios de agregación. Algunas, como PlumX y Dimensions, ya están integradas en repositorios institucionales mediante sistemas automatizados basados en identificadores persistentes como el DOI, lo que posibilita un seguimiento continuo y actualizado de la visibilidad digital de los conjuntos de datos.

En el marco de este estudio, no se han detectado señales alométricas relevantes asociadas a los conjuntos de datos analizados, lo que evidencia una proyección todavía limitada de estos recursos en los entornos digitales de comunicación científica y social. Sin embargo, se han identificado algunas citaciones académicas formales a ciertos *datasets*, lo que podría interpretarse como un primer indicio de atención y posible reutilización en el ámbito investigador. Este resultado refuerza la idea de que el acceso abierto, aunque necesario, no es suficiente por sí solo para asegurar la visibilidad ni el impacto efectivo de los datos depositados.

En consecuencia, si se aspira a que los conjuntos de datos generen valor —ya sea en forma de impacto académico, social o institucional—, resulta imprescindible avanzar hacia estrategias activas de comunicación, contextualización y diseminación. La publicación en repositorios constituye una condición necesaria, pero no suficiente. El impacto no se deriva del archivo, sino del uso, y este solo puede materializarse si los datos están adecuadamente descritos, enlazados a publicaciones, difundidos en los canales pertinentes y visibilizados ante sus públicos potenciales. Transformar los datos en conocimiento útil exige movimiento, tanto desde quienes los generan como desde quienes los gestionan.

Dada la escasa presencia alométrica observada en los conjuntos de datos analizados a nivel regional, se ha optado por incorporar estudios de caso internacionales como estrategia comparativa. Estos ejemplos permiten identificar qué factores favorecen una mayor circulación y visibilidad pública de los *datasets*, así como su potencial apropiación social.

Uno de los casos más ilustrativos es el *World Mortality Dataset*<sup>5</sup>, creado para monitorizar el exceso de mortalidad durante la pandemia de COVID-19 y publicado en enero de 2021.

5 Karlinsky, A. y Kobak, D. (2021). *The World Mortality Dataset: Tracking excess mortality across countries during the COVID-19 pandemic*. Epidemiology. <https://doi.org/10.1101/2021.01.27.21250604>

Este conjunto de datos figura en Dimensions, vinculado a una publicación científica de referencia y asociado a miles de interacciones altmétricas. Su impacto se traduce en más de 2.000 menciones, que incluyen cobertura mediática, presencia en documentos de políticas públicas, blogs especializados, redes sociales y Wikipedia. La urgencia del tema, su relevancia global y su accesibilidad facilitaron su rápida difusión y reutilización por parte de gobiernos, ciudadanía y comunidad científica. Es un ejemplo claro de cómo la combinación de pertinencia temática, calidad técnica y estrategias de comunicación eficaces puede convertir un *dataset* en un recurso con verdadero impacto transformador.

Otro caso relevante, esta vez fuera del ámbito sanitario, es el informe de la Oficina Nacional de Estadística del Reino Unido<sup>6</sup> sobre los precios de la vivienda en pequeñas áreas de Inglaterra y Gales. Publicado en junio de 2023 y recogido en la plataforma Overton, el documento se basa de forma explícita en múltiples conjuntos de datos abiertos. En este caso, los datos no solo alimentan el análisis, sino que fundamentan decisiones en áreas como la planificación urbana, la fiscalidad o el acceso a la vivienda.

Este tipo de reutilización muestra el valor estratégico de los datos abiertos para el diseño y el seguimiento de políticas públicas basadas en evidencia. Favorece la transparencia institucional, fortalece la gobernanza participativa y refuerza el vínculo entre ciencia y toma de decisiones. Estas prácticas, lejos de ser anecdóticas, deberían consolidarse como estándar dentro de las agendas de ciencia abierta con orientación social.

Por último, destaca también el caso de un *dataset* sobre recomendaciones cinematográficas a partir de Twitter, publicado junto a un artículo en *ACM Transactions on Intelligent Systems and Technology*<sup>7</sup> y registrado en Lens.org. Además de generar impacto académico a través de citas, este conjunto de datos ha sido referenciado en al menos tres patentes internacionales, evidenciando un uso real en procesos de innovación tecnológica y transferencia de conocimiento.

Este tipo de reutilización evidencia cómo un conjunto de datos bien documentado, accesible y contextualizado puede ir más allá del entorno académico y ser aprovechado por el sector privado o industrial, generando valor tanto económico como social. También muestra que la apertura y la visibilidad de los datos no solo favorecen su circulación, sino que amplían su impacto al conectarlos con procesos de innovación, desarrollo tecnológico y competitividad. Casos como este invitan a replantear los modelos de evaluación científica, incorporando indicadores que reflejen la contribución de los datos a la transferencia de conocimiento y a la actividad productiva.

Los tres ejemplos analizados son representativos del potencial transformador que pueden alcanzar los datos abiertos cuando se integran con éxito en dinámicas sociales, institucionales o económicas. Aunque se trata de experiencias concretas y muy contextualizadas, permiten

6 Office for National Statistics. (2023, 20 de septiembre). *Median house prices for administrative geographies: HPSSA dataset 9*. <https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianhousepriceforationalandsubnationalgeographiesquarterlyrollingyearhpssadataset09>

7 Dooms, S., Bellogín, A., De Pessemier, T. y Martens, L. (2016). A framework for dataset benchmarking and its application to a new movie rating dataset. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3), Article 41, 1-28. <https://doi.org/10.1145/2751565>

identificar condiciones clave para maximizar ese impacto: accesibilidad técnica, documentación clara, relevancia temática, buenas estrategias de difusión y capacidad para responder a necesidades reales del entorno.

En conjunto, estos casos muestran de forma clara que los datos abiertos, cuando se comunican bien, se comprenden y se reutilizan en contextos estratégicos, pueden convertirse en herramientas valiosas para la toma de decisiones públicas, la innovación intersectorial y la creación de valor social. Su análisis refuerza la necesidad de avanzar hacia marcos de evaluación que reconozcan este tipo de impactos emergentes y de fortalecer políticas que fomenten una circulación activa y significativa de los resultados científicos en forma de datos.

#### 4. CONCLUSIONES

A partir del análisis realizado, puede afirmarse que los datos en acceso abierto —especialmente aquellos depositados en repositorios institucionales— constituyen una vía incipiente pero estratégica para avanzar hacia modelos de ciencia más abiertos, responsables y orientados al impacto social. Sin embargo, los resultados evidencian que, pese a los avances en infraestructura y normativa, persisten importantes limitaciones en la normalización de los registros, la visibilidad efectiva de los datos y su reutilización, tanto en entornos académicos como no académicos.

En este contexto, se destacan tres ideas clave que orientan el debate y las acciones futuras:

Primero, es fundamental consolidar una cultura institucional y técnica que esté realmente alineada con los principios FAIR. No basta con permitir el depósito de datos; se necesitan mecanismos coordinados que aseguren su correcta identificación, descripción, contextualización y recuperación. Las diferencias observadas entre repositorios del Consorcio BUCLE reflejan una clara asimetría que afecta a la trazabilidad y la visibilidad de los datos, y que podría abordarse mediante políticas regionales de armonización y formación especializada.

Segundo, resulta imprescindible incorporar la dimensión comunicativa como parte estructural del ciclo de vida de los datos. El impacto —ya sea académico o social— no se produce simplemente por archivar un conjunto de datos, sino por su capacidad para circular, ser entendido y utilizado por otros actores. La ausencia de indicadores altmétricos o de citación no debe interpretarse como un fallo de las métricas, sino como una señal de que los procesos de comunicación y promoción de los datos aún son periféricos en el ecosistema científico.

Tercero, es necesario superar la lógica del cumplimiento mínimo y avanzar hacia una visión transformadora. El valor de los datos no se agota en el momento en que se depositan; al contrario, comienza ahí. Es clave que circulen, se comprendan y se usen. Que dejen de ser archivos estáticos y se conviertan en conocimiento activo. Solo así será posible una ciencia más conectada con su entorno y con mayor utilidad social.

A partir de este trabajo, se han iniciado nuevas líneas de investigación con el objetivo de profundizar en el papel de los datos abiertos como vectores de impacto:

- Análisis de casos internacionales con alto impacto público, para identificar factores que favorecen la circulación de *datasets*: calidad documental, relevancia temática,

políticas de comunicación y coordinación institucional. Esta comparación busca extraer buenas prácticas aplicables a otros contextos.

- Estudios cualitativos con agentes clave, como personal técnico de bibliotecas, oficinas de datos o servicios de apoyo a la investigación. Ya en fase de diseño, esta línea permitirá explorar las dimensiones institucionales que condicionan la publicación y la valorización de los datos.
- Seguimiento longitudinal de indicadores de uso, como visualizaciones, descargas y menciones digitales, analizados en relación con variables estructurales (por ejemplo, el uso de DOI, la existencia de *datapaper* o la temática del conjunto). El objetivo es entender los patrones y ritmos de circulación de los datos.
- Investigación sobre estrategias de comunicación y apropiación social, centrada en experiencias de cocreación, visualización narrativa o difusión en espacios no especializados. Esta línea busca evaluar el potencial de enfoques participativos para activar y hacer comprensibles los datos.
- Desarrollo de marcos evaluativos mixtos, que integren indicadores cuantitativos y cualitativos para medir el impacto de los datos de forma más completa y contextualizada. Este enfoque pretende superar las limitaciones de los modelos métricos tradicionales.

En conjunto, estas líneas conforman una agenda de investigación en expansión, orientada no solo a describir el estado actual de los datos abiertos, sino también a impulsar su visibilidad, uso y conexión con los desafíos sociales contemporáneos. Lejos de representar una fase final del proceso científico, los *datasets* son una interfaz activa entre ciencia y sociedad. Estudiar cómo circulan, se reapropian y generan valor es una tarea prioritaria, y ya en marcha.

## REFERENCIAS BIBLIOGRÁFICAS

- Anagnostou, P., Capocasa, M., Milia, N. y Bisol, G. D. (2013). Research data sharing: Lessons from forensic genetics. *Forensic Science International: Genetics*, 7(6), e117-e119. <https://doi.org/10.1016/j.fsigen.2013.07.012>
- Burgelman, J.-C., Pasco, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K. y Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Frontiers in Big Data*, 2(43). <https://doi.org/10.3389/fdata.2019.00043>
- Cole, N. L., Kormann, E., Klebel, T., Apartis, S. y Ross-Hellauer, T. (2024). The societal impact of Open Science: A scoping review. *Royal Society Open Science*, 11(6), 240286. <https://doi.org/10.1098/rsos.240286>
- Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M. y Clark, T. (2018). A data citation roadmap for scientific publishers. *Scientific Data*, 5(1), 180259. <https://doi.org/10.1038/sdata.2018.259>
- Fecher, B. y Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2272036>

- García Blázquez, P. I. (2024). Bibliometrías tradicionales y alométricas: un estudio aplicado a los Institutos Universitarios de Investigación de la Universidad de Salamanca [Trabajo de fin de máster, Universidad de Salamanca]. Gredos. <http://hdl.handle.net/10366/159447>
- Khan, N., Thelwall, M. y Kousha, K. (2021). Measuring the impact of biodiversity datasets: Data reuse, citations and altmetrics. *Scientometrics*, 126(4), 3621-3639. <https://doi.org/10.1007/s11192-021-03890-6>
- Penfield, T., Baker, M. J., Scoble, R. y Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1), 21-32. <https://doi.org/10.1093/reseval/rvt021>
- Peters, I., Kraker, P., Lex, E., Gumpenberger, C. y Gorraiz, J. (2016). Research data explored: An extended analysis of citations and altmetrics. *Scientometrics*, 107(2), 723-744. <https://doi.org/10.1007/s11192-016-1887-4>
- Quarati, A. y Raffaghelli, J. E. (2022). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, 48(4), 423-448. <https://doi.org/10.1177/0165551520961048>
- Thelwall, M. (2018). Using Altmetrics to Support Research Evaluation. En M. Erdt, A. Sesagiri Raamkumar, E. Rasmussen e Y.-L. Theng (Eds.), *Altmetrics for Research Outputs Measurement and Scholarly Information Management*, vol. 856, pp. 11-28. Springer Singapore. [https://doi.org/10.1007/978-981-13-1053-9\\_2](https://doi.org/10.1007/978-981-13-1053-9_2)




## LA CURACIÓN DE DATOS COMO METODOLOGÍA DE INVESTIGACIÓN

### *DATA CURATION AS RESEARCH METHODOLOGY*

Valeria MONTOYA-RONCANCIO<sup>1</sup> e Inmaculada BRAVO GARCÍA<sup>2</sup>

<sup>1</sup>*Universidad de Salamanca. España*

[valeriamontoya@usal.es](mailto:valeriamontoya@usal.es)

 <https://orcid.org/0000-0001-7357-3326>

<sup>2</sup>*Universidad de Salamanca. España*

[inma@usal.es](mailto:inma@usal.es)

 <https://orcid.org/0000-0003-3388-0545>

**RESUMEN:** En el contexto actual de creciente dependencia de los datos abiertos y reutilizables en la evaluación de la ciencia, la curación de datos se presenta como una metodología esencial en el campo de la Documentación. En esta investigación, se propone una metodología que combina técnicas de recuperación, normalización y análisis de datos, centradas en la verificación de registros provenientes de plataformas bibliográficas de datos como Web of Science (WoS) y Scopus utilizando identificadores persistentes como DOI, ORCID, ResearcherID, ScopusID e ISSN, con el fin de garantizar la fiabilidad del estudio. Este enfoque permite identificar inconsistencias, duplicidades o ausencias de información relevante en los perfiles de autor y publicaciones científicas. Los resultados muestran la producción de la Universidad de Salamanca, como estudio de caso, donde se identificaron perfiles de autores con datos que requerían corrección en ambas plataformas y en el portal de producción científica institucional. Estas incidencias evidencian la necesidad de procesos sistemáticos de curación en la gestión documental, destacando el uso de este modelo para mejorar la calidad de los sistemas de información científica. Su aplicación en el ámbito de la Documentación favorece la generación de métricas confiables y la toma de decisiones basada en información verificada y reutilizable.

**PALABRAS CLAVE:** curación de datos; evaluación científica; metadatos; identificadores persistentes; Web of Science; Scopus.

**ABSTRACT:** In the current context of increasing reliance on open and reusable data in the evaluation of science, data curation is presented as an essential methodology in the field

of Documentation. In this research, we propose a methodology that combines data retrieval, normalization and analysis techniques, focused on the verification of records coming from bibliographic data platforms such as Web of Science (WoS) and Scopus using persistent identifiers such as DOI, ORCID, ResearcherID, ScopusID and ISSN, in order to ensure the reliability of the study. This approach allows identifying inconsistencies, duplicities or absence of relevant information in author profiles and scientific publications. The results show the production of the University of Salamanca, as a case study, where author profiles were identified with data that required correction in both platforms and in the institutional scientific production portal. These incidences show the need for systematic curation processes in document management, highlighting the use of this model to improve the quality of scientific information systems. Its application in the field of Documentation favors the generation of reliable metrics and decision making based on verified and reusable information.

KEYWORDS: data curation; scientific evaluation; metadata; persistent identifiers; Web of Science; Scopus.

## 1. INTRODUCCIÓN

La creciente disponibilidad de información científica y la transformación digital del conocimiento han situado a los datos en el centro de procesos de investigación, gestión, evaluación y reutilización de contenidos. En este contexto, la curación de datos se presenta como una metodología clave que permite garantizar la consistencia, la calidad y la veracidad de la información proveniente de plataformas bibliográficas y bibliométricas que se encuentra depositada en los portales de investigación de las universidades.

Este capítulo ofrece una aproximación a la curación de datos desde el ámbito de las Ciencias de la Documentación, abordando su definición, su aplicación en la localización y tratamiento de datos académicos, así como su utilidad para la evaluación científica. Se examinan las principales herramientas de referencia, como Web of Science y Scopus junto con el papel de los identificadores persistentes en la recuperación masiva de registros bibliográficos. La propuesta metodológica se orienta a optimizar el uso de estos datos contribuyendo a una gestión documental más transparente, eficiente y alineada con los estándares de calidad a los que se enfrenta la comunidad científica.

### 1.1. DEFINICIÓN Y FUNDAMENTOS DE LA CURACIÓN DE DATOS

La curación de datos se entiende como el proceso que permite identificar las fuentes de datos, organizarlas, descargarlas, limpiarlas y prepararlas para la creación de análisis y extracción de información. Estas actividades están destinadas para que los datos de investigación sean comprensibles y reutilizables e incluyen la documentación, estandarización, formateo y asociación de metadatos relevantes para promover la reutilización interdisciplinaria de los datos científicos (Minamiyama, Takeda, Hayashi, Asaoka y Yamaji, 2024).

La curación de datos también se puede definir más específicamente como «las verificaciones y acciones realizadas por los curadores con el fin de asegurar que el conjunto de datos esté estructurado y documentado de la manera más completa posible y de acuerdo con las mejores prácticas» (SciELO, 2023). A su vez, la curación de datos se describe como una actividad de gestión que implica el desarrollo de infraestructuras lógicas y físicas que permite recolectar, indexar, almacenar y facilitar la consulta de los datos para análisis posteriores. Este enfoque evidencia la importancia de las infraestructuras sociotécnicas y de la gestión para acceder a los datos (Chua *et al.*, 2022; citado por Parmiggiani *et al.*, 2024).

Según Choi y Xin (2021), la curación de datos es el proceso que permite gestionar datos para que estén disponibles para su reutilización y preservación, permitiendo usos que cumplan con los requisitos FAIR (Findable, Accessible, Interoperable, and Reusable). Los principios FAIR establecen que los datos científicos deben ser localizables, accesibles, interoperables y reutilizables, con el fin de maximizar su visibilidad, integrabilidad y valor en la investigación. Estos autores destacan la necesidad de la curación en el ciclo de vida de la investigación, asegurando que los datos sean descubiertos y reutilizados eficientemente.

Yakel (2007) define la curación de datos como la gestión activa y continua de los datos a lo largo de su ciclo de vida, con el fin de conservar su utilidad para la ciencia, la investigación y la educación a lo largo del tiempo, resaltando el alcance de la curación en el acceso y la preservación de datos útiles para la comunidad académica.

Estas definiciones reflejan la evolución del concepto de curación de datos en la literatura científica, mostrando el papel crucial que este proceso posee en la gestión, la preservación y la reutilización de datos en diversos contextos de investigación. La curación de datos implica diversas acciones para asegurar que los datos sean accesibles y útiles para futuras investigaciones.

El objetivo de este estudio es describir una estrategia de automatización para la curación de datos en el campo de la Documentación. Un campo científico en el que es necesario no solo almacenar o recuperar información; sino que estudia cómo se produce, organiza, conserva y difunde el conocimiento, y requiere métodos claros y sistemáticos para que sus resultados sean válidos y replicables. Actualmente, en esta rama del conocimiento los datos se tratan cada vez más desde una perspectiva técnica, mediante herramientas que permiten identificar metadatos incompletos, mejorar la interoperabilidad entre plataformas y verificar la persistencia de identificadores. La curación de datos se consolida así como una metodología clave para garantizar que los registros documentales sean fiables, estandarizados y útiles para la evaluación, la investigación y la toma de decisiones informadas (Peng y Wyborn, 2022).

El modelo de curación de datos que se presenta en esta investigación se basa en la extracción y la estructuración de datos abiertos de carácter de autoría y bibliográfico. La propuesta parte de la localización de datos de interés para la investigación relativos a publicaciones y autores mediante el tratamiento de datos de identificadores persistentes procedentes de plataformas bibliográficas y bibliométricas como Web of Science y Scopus.

## 1.2. LOCALIZACIÓN DE DATOS EN PLATAFORMAS BIBLIOGRÁFICAS Y BIBLIOMÉTRICAS

Web of Science (WoS), operada por Clarivate, es una plataforma de información científica que proporciona acceso a publicaciones académicas. Su principal función es facilitar la búsqueda, el análisis y la evaluación de literatura científica y técnica de alta calidad, cubriendo una amplia gama de disciplinas. Esta base de datos bibliométrica es clave para calcular indicadores como el Índice H y el Factor de Impacto (FI). Se caracteriza por su rigurosidad en la selectividad de revistas científicas y en sus índices especializados como el Science Citation Index o el Social Sciences Citation Index.

Scopus, gestionado por Elsevier, es una base de datos multidisciplinar que ofrece una amplia literatura académica, científica y técnica. Contiene información bibliográfica y de citación de artículos publicados en revistas, libros y actas de congresos. Es especialmente valorada por su capacidad de generar métricas de impacto y colaboración científica. Scopus permite identificar tendencias de investigación, redes de coautoría y autores destacados. Ambas plataformas, WoS y Scopus, son esenciales para el análisis y la medición de la actividad investigadora y constituyen las principales fuentes de referencia para la extracción de indicadores bibliométricos empleados en la evaluación de la ciencia. Además, permiten calcular métricas como el Índice H, el número de citas, el impacto y la visibilidad de las publicaciones.

## 1.3. RECUPERACIÓN MASIVA DE REGISTROS BIBLIOGRÁFICOS

Las interfaces de programación de aplicaciones (API, por sus siglas en inglés) constituyen un conjunto estandarizado de métodos que permiten acceder a los datos ofrecidos por plataformas tales como Scopus y Web of Science, de forma controlada, eficiente y con campos y formatos predefinidos por cada proveedor (Torres-Salinas *et al.*, 2022; Vélez-Estévez *et al.*, 2023). Esta vía facilita la recuperación masiva y precisa de registros bibliográficos, así como su integración en sistemas propios de análisis documental. Por su parte, el *web scraping* es una técnica informática que permite extraer información directamente de páginas web en las que no se ofrece acceso a través de una API oficial, mediante el análisis del código HTML y la simulación de la navegación humana.

La API de Web of Science, desarrollada por Clarivate, permite acceder a los datos indexados en la Web of Science Core Collection. Esta interfaz está diseñada para usuarios institucionales que cuenten con una suscripción activa y una clave de autenticación. A través de esta API, es posible recuperar metadatos bibliográficos detallados, información sobre autores, afiliaciones, recuentos de citas, referencias bibliográficas y otros elementos clave para el análisis bibliométrico. Por su parte, la API de Scopus, gestionada por Elsevier, proporciona acceso estructurado a la base de datos bibliográfica Scopus mediante una clave de API asociada a una suscripción institucional. Esta interfaz ofrece funcionalidades similares a las de Web of Science, como la búsqueda y la recuperación de publicaciones, citas, referencias y metadatos de autores y afiliaciones.

Para el caso de DOI, la IDF (International DOI Foundation) gestiona un sistema de resolución basado en el Handle System y proporciona una API en <https://doi.org/> que actúa como un sistema de resolución de identificadores DOI, redirigiendo cada DOI a la ubicación actual del recurso digital al que hace referencia. Si bien la IDF se estableció en 1997, la infraestructura para solicitar y asignar DOI no estuvo operativa hasta 2000. CrossRef fue el primer registro de DOI especializado en publicaciones académicas, facilitando la adopción masiva del sistema por revistas científicas y editoriales.

Este estudio describe una estrategia para la recuperación de datos a través de API, llevando a cabo la extracción, la transformación y la carga de los datos que, convenientemente estructurados y relacionados, ofrecerán información para la realización de otros estudios e investigaciones. La curación de datos permitirá detectar duplicidades, errores e inconsistencias garantizando que los datos consultados sean precisos, útiles y sirvan como base para la toma de decisiones, implementación de políticas y gestión documental. En la metodología que se presenta, se evidenciaron las herramientas utilizadas para la recuperación de información, el tratamiento de los datos y la exportación de los mismos, lo que facilitará la obtención de indicadores relevantes, así como la identificación de incidencias, al consultar algunos identificadores persistentes.

#### 1.4. IDENTIFICADORES PERSISTENTES COMO ELEMENTOS CLAVE

En el ecosistema de la información científica, los identificadores persistentes desempeñan un papel clave para garantizar la trazabilidad, la desambiguación y la vinculación correcta de entidades como autores, instituciones y publicaciones. Los identificadores persistentes se emplean como el medio para localizar información de investigación de modo eficiente y precisa de autores con firmas parecidas.

Entre los identificadores más relevantes se encuentra el DOI (Digital Object Identifier). Un DOI (Digital Object Identifier) es un identificador único y persistente asignado a un objeto digital, como un artículo académico, para garantizar su localización y acceso a largo plazo. Su sintaxis siempre consta de un prefijo numérico «10» seguido de un código numérico asignado al registrante (editor o publicador) y un sufijo único para cada objeto. La forma típica de un DOI es: 10.xxxx/xxxxxxxxx. La asignación de DOI a publicaciones comenzó en el año 2000, cuando CrossRef, una organización fundada por editores académicos, empezó a registrar identificadores DOI para artículos científicos y otros documentos.

Si bien la International DOI Foundation (IDF) se estableció en 1997, la infraestructura para solicitar y asignar DOIs no estuvo operativa hasta 2000. CrossRef fue el primer registrador de DOIs especializado en publicaciones académicas, facilitando la adopción masiva del sistema por revistas científicas y editoriales. Desde entonces, el uso del DOI se ha expandido a otros ámbitos, incluyendo *datasets*, reportes, tesis y hasta materiales audiovisuales.

En el ámbito de la identificación de autores, destaca el ORCID (Open Researcher and Contributor ID); es un identificador digital persistente lanzado en 2012 por ORCID. Es gestionado por el propio investigador y permite unificar su producción científica en distintas

plataformas, promoviendo la interoperabilidad y la correcta atribución de la actividad académica. ORCID es un identificador abierto que permite consolidar la producción científica de un investigador más allá de las variaciones nominales.

ResearcherID es un sistema de identificación de autores científicos. El sistema fue introducido en enero de 2008 por Thomson Reuters Corporation. Este identificador único tiene como objetivo solucionar el problema de la identificación de autores y la correcta atribución de obras dentro de su base de datos. En 2019 se integra con Publons y en 2022 Publons desaparece y se integra en Web of Science (Fundación Española para la Ciencia y la Tecnología, 2022).

Asimismo, ScopusID es un identificador único generado automáticamente por la base de datos bibliográfica Scopus, gestionada por Elsevier, que agrupa todas las publicaciones de un autor bajo un mismo perfil. Este sistema facilita la correcta identificación y la desambiguación de autores, especialmente en casos de homónimos o variaciones en la forma de escribir el nombre.

Ambos identificadores, ResearcherID y ScopusID, permiten consolidar métricas bibliométricas, como el número de citas y el Índice H, asociados a la producción científica de cada investigador. Los autores pueden solicitar la corrección o la fusión de sus perfiles para asegurar la precisión de los datos. Son herramientas fundamentales para la evaluación del rendimiento académico y el análisis de redes de colaboración.

A nivel institucional, el ROR (Research Organization Registry) actúa como identificador persistente para instituciones académicas, contribuyendo a mejorar la interoperabilidad de los datos de filiación.

Para identificar las revistas científicas se utiliza el ISSN. El ISSN (International Standard Serial Number) es un identificador numérico normalizado de ocho dígitos que se asigna a las publicaciones seriadas, como revistas científicas, boletines y anuarios, con el fin de facilitar su identificación unívoca a nivel internacional. Establecido por la norma ISO 3297, el ISSN no guarda relación con el contenido ni con la editorial, sino que distingue cada título de publicación seriada de forma independiente. El octavo dígito es un código de control que se calcula en función de un algoritmo Módulo 11, sobre la base de los 7 dígitos anteriores.

La gestión global del sistema ISSN está a cargo del Centro Internacional del ISSN, cuya plataforma principal es el sitio web ISSN.org. Desde allí, se coordina una red de centros nacionales e institucionales que se encargan de asignar y mantener los ISSN en sus respectivos países. Además, el sitio proporciona acceso a la base de datos ISSN Portal, un registro internacional donde es posible consultar información normalizada sobre publicaciones seriadas registradas, incluyendo detalles como título, editor, país, frecuencia y soporte. Este sistema cumple un papel esencial en la trazabilidad, la evaluación y la preservación del registro bibliográfico de la literatura seriada en el ámbito académico y editorial. El Centro Internacional del ISSN ha asignado más de 2.5 millones de ISSN. La base de datos del ISSN se actualiza constantemente y aumenta aproximadamente entre 50.000 y 70.000 ISSN por año.

Cuando una publicación está disponible en más de un soporte, por ejemplo, en formato impreso y electrónico, se le asignan dos ISSN diferentes: el ISSN para la versión impresa y el e-ISSN (o ISSN electrónico) para la versión digital. Esta distinción es fundamental en los procesos de catalogación, citación, indexación en bases de datos y evaluación de publicaciones,

ya que permite diferenciar claramente entre ediciones que pueden tener distintas fechas de publicación, paginación u otras características editoriales.

El número ISSN de enlace, o ISSN-L, es un número ISSN específico que reúne las ediciones en diferentes soportes (impreso, electrónico) de una misma publicación seriada (cada edición tiene su propio número ISSN). La tabla de números ISSN-L es un archivo con el cual los administradores de bases de datos o de catálogos pueden procesar con mayor eficacia los números ISSN presentes en sus fuentes, debido a que el número ISSN-L se utiliza como clave de enlace entre las diferentes ediciones identificadas por sus números ISSN específicos (ISSN International Standard Serial Number, s. f.).

Tal como señala Paloma Marín-Arraiza (2022), estos identificadores persistentes no solo favorecen la desambiguación en contextos bibliométricos complejos, sino que «forman parte de una infraestructura crítica para una ciencia abierta, conectada e interoperable», permitiendo el desarrollo de servicios avanzados en los sistemas de información científica. En este contexto, el análisis de la producción científica requiere no solo de identificadores normalizados, sino de herramientas que reflejen su impacto y visibilidad.

## 1.5. ESPAÑA Y SUS PRINCIPALES HERRAMIENTAS DE REFERENCIAS

En el caso de España, Web of Science y Scopus son las principales herramientas de referencia empleadas por las agencias evaluadoras como la FECYT (Fundación Española para la Ciencia y la Tecnología) y ANECA (Agencia Nacional de Evaluación de calidad y Acreditación) para valorar la calidad y el impacto para la producción científica de investigadores, revistas y universidades.

La FECYT, en el ámbito de la evaluación científica, no solo gestiona el acceso a las bases de datos como WoS y Scopus, sino que coordina los procesos de acreditación de revistas científicas españolas garantizando su calidad y visibilidad internacional. Este reconocimiento se otorga a través del Sello de Calidad FECYT, donde el impacto y la indexación en WoS y Scopus contribuyen positivamente en la evaluación de las revistas (FECYT, 2025). La ANECA es la entidad responsable de la acreditación académica, la medición de la calidad del profesorado universitario, la mejora institucional, la evaluación de sexenios y de los títulos oficiales en España. Para esta evaluación utiliza también criterios bibliométricos e indicadores de calidad basados en la calidad, la relevancia y el impacto de las contribuciones científicas en bases de datos como WoS y Scopus (ANECA, 2024).

Este estudio parte de un planteamiento sobre la falabilidad de la producción científica difundida en plataformas bibliográficas como son Web of Science y Scopus. En este sentido, Harari (2024) señala que «rechazar la fantasía de la infalibilidad, permite construir una red de información que considera que el error es inevitable». Esta perspectiva refuerza la necesidad de contar con una metodología rigurosa que permita el tratamiento de datos fiables, estructurados y persistentes, estableciendo un puente entre los requisitos institucionales de calidad, la gestión técnica de la información y la toma de decisiones fundamentada y transparente.

En este contexto, la curación de datos se convierte en una herramienta metodológica fundamental para acceder a los datos de una forma más eficaz y facilitar su reutilización. Esta metodología permite optimizar el uso de los datos extraídos en plataformas como Web of Science y Scopus, especialmente en el análisis científico y en los procesos de evaluación académica. Asimismo, al visibilizar y depurar estos datos, se pueden identificar incidencias que influyen directamente en la medición de la producción científica de los investigadores y, por tanto, en su impacto real dentro del ecosistema científico.

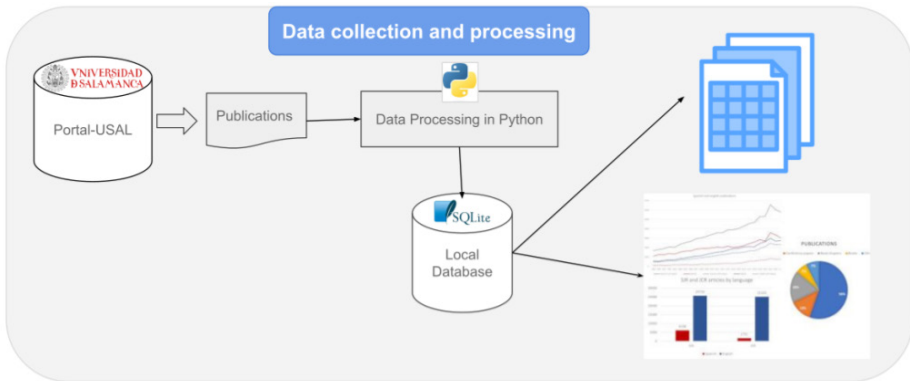
## 2. METODOLOGÍA

La metodología parte de una aproximación teórica y de un estudio de caso, planteado como estrategia de un caso real de curación de datos de investigación. La primera metodología empleada fue la investigación bibliográfica, mediante la que se realizó el contexto teórico de la curación de datos de investigación y los aspectos relevantes tratados en este estudio: extracción, transformación y estructuración de datos, métodos de gestión de datos y plataformas bibliográficas de referencia. El estado de la cuestión priorizará la descripción de las técnicas de gestión de datos más útiles, como son el uso de API, el *web scraping* o extracción de datos automatizada desde la web y los identificadores persistentes.

La segunda metodología es el estudio de caso, donde se presenta la estrategia práctica de curación de datos de aplicación en la investigación. Se introduce el modelo ETL como proceso de integración y carga de datos para el cálculo y el análisis de la información consultada. El procedimiento ETL consiste en Extract (Extracción), Transform (Transformación) y Load (Cargar). Por lo que se extraen los datos de sistemas heredados, se depuran para mejorar su calidad y establecer una coherencia y, por último, se introducen los datos en una base de datos de destino para su análisis (García García-Doncel, 2023).

En la arquitectura de procesamiento de datos bibliográficos, el modelo ETL constituye una metodología fundamental para sistematizar la curación de grandes volúmenes de información procedente de fuentes heterogéneas. Esta estrategia se estructura en tres fases diferenciadas pero interdependientes. En primer lugar, la fase de extracción (Extract) implica la recuperación de registros bibliográficos desde diferentes orígenes. Posteriormente, en la etapa de transformación (Transform), los datos extraídos son sometidos a procesos de depuración y estandarización que incluyen la eliminación de duplicados, la corrección de errores tipográficos y la normalización de los campos recuperados. Esta fase es crucial para garantizar la calidad, la coherencia y la comparabilidad de los datos. Finalmente, la etapa de carga (Load) consiste en la incorporación de los datos transformados a un sistema de almacenamiento o base de datos centralizada, optimizada para su posterior análisis bibliométrico, visualización o explotación. Este proceso se realizó tomando como estudio de caso el Portal de investigación de la Universidad de Salamanca y se puede observar en la Figura 1.

Figura 1. Modelo ETL



Fuente: Elaboración propia.

La curación de datos aplicada a este portal de investigación se llevó a cabo a través de las siguientes seis fases: recopilación de DOI, extracción de datos mediante API, obtención del listado de revistas científicas, análisis de identificadores de revistas (ISSN y EISSN), análisis de identificadores de autores y procesamiento y almacenamiento de datos.

## 2.1. PRIMERA FASE: RECOPIACIÓN DE DATOS DEL PORTAL DE PRODUCCIÓN CIENTÍFICA DE LA UNIVERSIDAD

En esta primera fase se realizó la compilación de los investigadores de la universidad y de las publicaciones con DOI correspondientes a cada autor. Esta información fue consolidada desde bases de datos internas o información proporcionada por los participantes.

Se decidió utilizar únicamente las publicaciones con DOI informado, para partir de un dato unívoco y así poder recuperar y comparar la misma publicación en diferentes fuentes, evitando generar errores derivados de la posible ambigüedad que se produciría de otra manera; de igual modo, los autores seleccionados para el análisis fueron los autores con al menos una publicación con DOI.

En esta fase, además de extraer las publicaciones con DOI, se desarrolló una depuración de datos, excluyendo las publicaciones afectadas para evitar propagar errores. Así mismo, se generaron informes con los tres errores analizados para que el personal técnico pudiera proceder a su resolución.

Se verificaron los siguientes posibles errores:

- DOI duplicados: es decir el mismo DOI hace referencia a dos o más publicaciones diferentes en el portal de producción científica. Lo cual no es coherente por la propia definición y funcionalidad del DOI que se ha descrito.

- DOI mal formados: no cumplen con el formato estandarizado de un DOI, no se ajusta a esta sintaxis: 10.xxxx/xxxxxxxxx.
- DOI no activos: identificadores que, pese a tener una sintaxis válida, no conducen a ninguna publicación accesible a través del sistema de resolución DOI. Adicionalmente, se realiza un análisis de todos los DOI utilizando la API de DOI.ORG.

## 2.2. SEGUNDA FASE: EXTRACCIÓN DE DATOS MEDIANTE API

En esta segunda fase se extrae la información necesaria de las API de WoS y Scopus y se almacena en una base de datos local para su posterior tratamiento.

Las consultas a estas API pueden realizarse utilizando diversos parámetros. En este estudio, se ejecutaron búsquedas por DOI y por autor, utilizando el ReseracherID y el ScopusID respectivamente. Los metadatos que se recuperaron fueron, principalmente, el DOI de la publicación, los nombres de los autores, sus identificadores en la plataforma y sus ORCID.

En Web of Science, se emplearon solicitudes HTTP con *requests* para extraer información equivalente a través de su API REST. En Scopus, se utilizó la Biblioteca Pybliometrics para interactuar con la API REST de Scopus y recuperar información en formato JSON, incluyendo identificadores de autores (ScopusID) y metadatos de publicaciones (Rose y Kirchin, 2024) (Figura 2).

Figura 2. Extracción de datos



Fuente: Elaboración propia.

## 2.3. TERCERA FASE: OBTENCIÓN DEL LISTADO DE REVISTAS CIENTÍFICAS

Para poder hacer una comparativa de revistas indexadas en cada plataforma se requería de un listado completo de las revistas con sus respectivos ISSN para realizar la comparativa de manera precisa. En Scopus, se descargó el listado completo de revistas indexadas proporcionado en su sitio web; en Web of Science, se emplearon técnicas de *web scraping* con selenium y BeautifulSoup4 para extraer el listado completo desde su sitio web. Se trataron los datos extraídos, unificándolos y cargándose en la base de datos local SQLite para su posterior tratamiento.

Ambos procesos se ejecutaron en octubre de 2024.

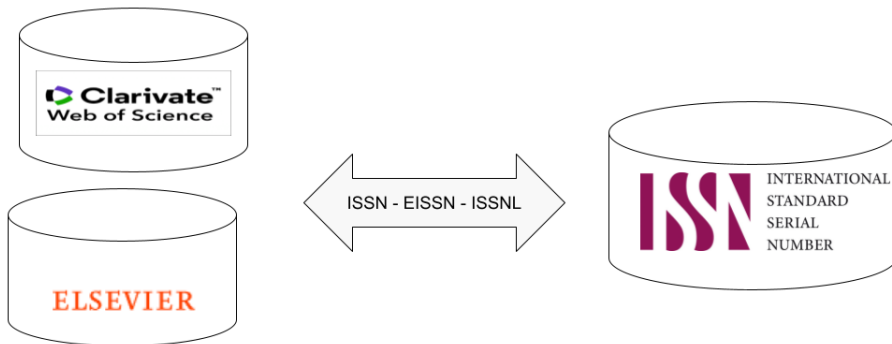
#### 2.4. CUARTA FASE: ANÁLISIS DE IDENTIFICADORES DE REVISTAS (ISSN Y EISSN)

En esta fase se realizó una depuración de los datos sobre revistas obtenidos en la fase anterior. Primero, se comprobó la existencia de al menos un código ISSN en cada revista. Seguidamente, se comprobó la validez de los códigos ISSN, verificando que el octavo dígito, correspondiente al dígito de control, cumpliera correctamente su función según el algoritmo de verificación Módulo 11, aplicado sobre los siete dígitos numéricos previos. Para este proceso se realizó un script en Python para su procesamiento automático.

En el proceso de comparación de los listados de revistas de ambas plataformas se emplearon como claves principales los identificadores ISSN y EISSN. No obstante, se detectaron inconsistencias en la asignación de estos códigos: en algunos casos, un mismo identificador aparecía como ISSN en una base de datos y como EISSN en la otra. Para subsanar estas discrepancias y asegurar una comparación homogénea, se recurrió al uso del ISSN-L (Linking ISSN), que permite agrupar todas las versiones de una publicación seriada bajo un único identificador, lo que facilitó la normalización de los registros y una comparación fiable entre las dos fuentes. Se obtuvo la tabla completa de estos identificadores solicitándolo a través del portal oficial ISSN.org.

De esta manera, se analizó el ISSN-L de todas las revistas incluidas en Web of Science y Scopus, bajo el supuesto de que dicho identificador debería ser único para cada título, al vincular todas sus variantes de ISSN, tanto impresas como electrónicas. No obstante, se identificaron casos en los que una misma revista presentaba múltiples ISSN-L, lo que contradecía su función como identificador unificador (Figura 3). Ante esta inconsistencia, se adoptó un criterio operativo más flexible: se consideró que dos registros correspondían a la misma revista si coincidían sus ISSN o EISSN o si el ISSN de una base coincidía con el EISSN de la otra.

Figura 3. Identificadores de revistas



Fuente: Elaboración propia.

## 2.5. QUINTA FASE: ANÁLISIS DE IDENTIFICADORES DE AUTORES

Se examinaron los identificadores persistentes de autores, como ResearcherID (de Web of Science) y ScopusID de Scopus, realizando una comparación sistemática con los identificadores presentes en el portal de producción científica en cada publicación. El análisis detectó casos de perfiles duplicados y asignación incorrecta de autorías de publicaciones.

Otras incidencias menores que se detectaron fueron las siguientes: error de orden, es decir, el autor no se encuentra en el mismo orden de firma de la publicación en origen, esto es, en el portal institucional y en la plataforma analizada; error de orden excedido, lo cual ocurre cuando en origen el orden de firma supera al número de autores recuperados para una publicación a través de la API.

## 2.6. SEXTA FASE: PROCESAMIENTO Y ALMACENAMIENTO DE DATOS

Los datos extraídos fueron procesados y almacenados en una base de datos local SQLite, garantizando su portabilidad y facilidad para futuros análisis; también se crearon informes en formato de hojas de cálculo con enlaces a las fuentes originales de cada dato para facilitar la verificación y la corrección manual de las incoherencias detectadas. Estos informes se generaron de manera automatizada utilizando la librería 'xlsxwriter' de Python.

# 3. RESULTADOS

Una vez aplicada la metodología, se hallaron resultados de investigación de dos tipos: teóricos y de aplicación práctica. Los resultados teóricos consistieron en la descripción de las técnicas de mayor utilidad para la curación de datos de investigación, especialmente en lo relativo al uso de las API de las bases de datos, así como a los métodos de extracción de datos desde la web o *web scraping*, los identificadores persistentes y principales herramientas de referencia (Khder, 2021).

Los resultados prácticos se orientan hacia la aplicación de una estrategia útil para la curación de datos de investigación; su procesamiento, estructuración y reutilización para la obtención de indicadores, y como metodología para la demostración de hipótesis. Se facilitan informes en formato excel a los técnicos responsables de la curación de datos.

### 3.1. RESULTADOS OBTENIDOS EN LA RECOPIACIÓN DE DATOS DEL PORTAL DE PRODUCCIÓN CIENTÍFICA DE LA UNIVERSIDAD

En la primera fase, al extraer los datos del portal institucional de las publicaciones con DOI y analizarlos, se obtuvo que el 76,9 % están presentes en WoS y el 87,9 % están presentes en Scopus. Sobre los errores en origen detectados, el 0,7 % son DOI duplicados, es decir, que

el mismo DOI hace referencia a dos publicaciones diferentes dentro del portal; el 0,6 % son DOI no activos, es decir, que no remiten a una publicación disponible, y únicamente un 0,01 % con errores de formato (Tabla 2).

Tabla 1. Publicaciones

DOI	En WoS	En Scopus	Duplicados	No activos	Error formato
50841	38849 (76,9 %)	44402 (87,9 %)	361 (0,7 %)	314 (0,6 %)	6 (0,01 %)

Fuente: Elaboración propia.

Con los datos de los errores detectados se generan informes en formato excel, incluyendo enlaces al portal de producción científica para facilitar su corrección: DOI duplicados en el portal de producción científica, DOI con errores de formato y DOI que no existe el documento al que deberían hacer referencia.

### 3.2. RESULTADOS OBTENIDOS EN EL ANÁLISIS DE IDENTIFICADORES DE REVISTAS (ISSN Y EISSN)

En la tercera fase, el análisis de identificadores de revistas científicas muestra los siguientes datos:

El número de revistas indexadas en WoS es de 21.273.

ISSN o EISSN no valido: 5.

Revistas sin ISSN ni EISSN: 0.

El número de revistas indexadas en Scopus es de 46.535.

ISSN o EISSN no valido: 33.

Revistas sin ISSN ni EISSN: 253.

El número de revistas indexadas por ambas plataformas es de 19.705.

En este estudio, al analizar los datos, se observó que revistas en Scopus y Web of Science tienen ambos identificadores, ISSN y E-ISSN, que deberían converger en el mismo ISSN-L, sin embargo, se encontraron 581 revistas con más de un ISSN-L. Se consultó a ISSN.org cómo deberíamos proceder para intentar resolver esta aparente inconsistencia en los datos. ISSN.org solicitó que se le enviara el listado completo y se les envió por correo electrónico el informe en formato excel con enlaces a la revista en WoS o Scopus y a su portal en febrero de 2025 (Tabla 2).

Tabla 2. Ejemplo de revistas con más de un ISSN-L

NOMBRE	ISSN	ISSN-L	E-ISSN	ISSN-L
ACTA CHIRURGICA BELGICA	0001-5458	0001-5458	2577-0160	1784-3421
Ornithology	0004-8038	0004-8038	2732-4613	2732-4613
BLUMEA	0006-5196	0006-5196	0373-4293	0373-4293
Ornithological Applications	0010-5422	0010-5422	2732-4621	2732-4621
Foreign Trade Review	0015-7325	0015-7325	0971-7625	0971-7625
Geofísica Internacional	0016-7169	0016-7169	2954-436X	2954-436X
Acta Cardiologica	0001-5385	0001-5385	0373-7934	0373-7934
German Journal of Agricultural Economics	0002-1121	0002-1121	2191-4028	2191-4028
Annales de Medecine Veterinaire	0003-4118	0003-4118	1781-3875	1781-3875

Fuente: Elaboración propia.

### 3.3. RESULTADOS OBTENIDOS EN EL ANÁLISIS DE IDENTIFICADORES DE AUTORES

El análisis de perfiles de autor en WoS evidenció que en el portal de origen estaba informado el ResearcherID del 48,3 % de los autores; se halló que el 37,1 % de los autores, aunque no disponen de este identificador en origen, sí tienen un perfil de autor en WoS.

Un dato relevante en este estudio es que del 30,8 % de los autores se recuperó más de un identificador ReseracherID en WoS.

El 69 % de los autores de la institución tiene algún dato que corregir, esto es, que no tenía informado el identificador en el portal de origen y sí en WoS, o se ha encontrado más de un perfil en WoS para dicho investigador, o se ha detectado alguna de las incidencias menores descritas en la metodología (Tabla 3).

Tabla 3. Perfiles de autor en WoS

Autores	Con RID en origen	Sin RID en origen y sí en WOS	Con más de un RID en WoS	Con algún dato que corregir
4628	2234 (48,3 %)	1719 (37,1 %)	1433 (30,8 %)	3193 (69,0 %)

Fuente: Elaboración propia.

El análisis de perfiles de autor en Scopus evidenció que en el portal de origen estaba informado el ScopusID del 76,8 % de los autores y se halló que el 10,7 % de los autores, aunque no dispone de este identificador en origen, sí tienen un perfil de autor en Scopus.

Un dato relevante en este estudio es que del 6.8 % de los autores se recuperó más de un identificador ScopusID en Scopus.

El 36,1 % de los autores de la institución tiene algún dato que corregir, esto es, que bien no tenía informado el identificador en el portal de origen y sí en Scopus, o se ha encontrado más de un perfil en Scopus para dicho investigador, o se han detectado alguna de las incidencias menores descritas en la metodología (Tabla 4).

Tabla 4. Perfiles de autor en Scopus

Autores	Con ScopusID en origen	SIN Scopus_ID en origen y sí en Scopus	Con más de un Scopus_ID en Scopus	Con algún dato que corregir
4628	3552 (76,8 %)	495 (10,7 %)	315 (6,8 %)	1669 (36,1 %)

Fuente: Elaboración propia.

### 3.4. RESULTADOS OBTENIDOS EN EL PROCESAMIENTO Y ALMACENAMIENTO DE DATOS

Finalmente, en la sexta fase se genera un informe, por cada plataforma analizada, con la información de los perfiles de autor con algún dato que corregir. En estos informes se agrupan en una fila las publicaciones de cada autor, con el mismo resultado de comparación entre los metadatos de origen, es decir, los publicados en el portal de producción científica y los hallados a través de la API en la plataforma analizada. Para facilitar la toma de decisión en cuanto a la corrección de la incidencia se incluyen enlaces a los perfiles de autor y a los documentos en el portal de origen, en ORCID, en WoS y en Scopus.

Figura 4. Informes

Informe de Scopus						
ID	Autor	Scopus_ID	S_Scopus_ID	S_autor	Clasificación origen/API-BD	Publicaciones
a1	autor1	scopusid1			OtrosErrores	1
						5
					ia_ID	3
a1	autor1	rid1	rid1	autor1	Correcto	3
a2	autor2	rid2	rid3	autor2	Incidencia_ID	6
a3	autor3				No existe en WOS	9
						9
a1	autor1	rid1			OtrosErrores	2
a4	autor4		rid4	autor4	Incidencia_ID	1

Fuente: Elaboración propia.

#### 4. DISCUSIÓN Y CONCLUSIONES

Los resultados de este estudio ponen en evidencia la importancia de la curación de datos como metodología eficaz para identificar inconsistencias y mejorar la calidad de la información científica. Las bases de datos Web of Science y Scopus, que tradicionalmente han sido consideradas referentes en la evaluación de la producción científica, presentan errores en sus registros. Sin embargo, son utilizadas para realizar análisis bibliométricos y rankings institucionales sin cuestionar la calidad y la consistencia de los datos que proporcionan.

En la revisión de las revistas indexadas por ambas bases de datos se observó que 19.705 están presentes en ambas plataformas, es decir, que el 90 % de las revistas indexadas en WoS también lo están en Scopus. Además, Scopus indexa más del doble de revistas científicas (46.535 en Scopus y 21.273 en WoS). Por este motivo, el porcentaje de publicaciones de la universidad en Scopus es mayor (87.9 % frente al 76.9 % en WoS). No obstante, el hecho de que WoS mantenga una proporción tan alta sugiere que, a pesar de su menor cobertura, sigue siendo la principal referencia para muchos investigadores. Esta preferencia puede explicarse por el valor que las publicaciones en WoS tienen en los procesos de evaluación científica gestionados por agencias como ANECA y FECYT, donde se otorgan puntuaciones decisivas para la obtención de acreditaciones, sexenios, financiación de proyectos, reconocimiento académico y evaluación de publicaciones.

Uno de los hallazgos más significativos fue que el 69 % de los autores de la institución analizada presentaban algún dato que corregir. Estas incidencias incluían que no tenían informado el identificador en el portal de origen y sí en WoS, o se había encontrado más de un perfil en WoS para dicho investigador o algunas incidencias menores.

Otro dato que llamó la atención fue la alta tasa de duplicidad de perfiles de autor en WoS: un 30,8 % de los investigadores tenían más de un perfil, frente al 6,8 % en Scopus. Esta diferencia puede deberse tanto a la mayor eficiencia de los algoritmos de detección y generación de autoría en Scopus como al hecho de que la universidad ha invertido más tiempo en depurar los datos provenientes de esta plataforma.

Por otro lado, se evidenció que el identificador de Scopus estaba ampliamente informado en el portal de producción científica de la universidad (77 % de los autores), mientras que solo el 48 % de los investigadores tienen registrado su ResearcherID de WoS. Esta diferencia se explica por el hecho de que Scopus fue la fuente inicial de datos en la construcción del portal institucional, lo que generó una integración más robusta desde el inicio.

Las incidencias halladas en esta investigación evidencian el impacto en la visibilidad y la evaluación de la producción científica de los investigadores, así como la falibilidad de los sistemas que sustentan las políticas de evaluación académica. La existencia de múltiples perfiles, la falta de normalización de los metadatos o la ausencia de identificadores puede ocasionar que una parte significativa del trabajo de los investigadores quede invisibilizada, tanto en portales institucionales como en las bases comerciales; comprometiendo la calidad de los análisis y la presencia en el ecosistema científico. En este sentido, cobra valor la frase de Harari (2024) cuando afirma que «rechazar la fantasía de la infalibilidad, permite construir una red de información que considera que el error es inevitable». Esta perspectiva

refuerza la necesidad de contar con una metodología rigurosa que permita el tratamiento de datos fiables, estructurados y persistentes, estableciendo un puente entre los requisitos institucionales de calidad, la gestión técnica de la información y la toma de decisiones fundamentada y transparente.

La curación de datos permite no solo estructurar grandes volúmenes de información, sino también establecer relaciones significativas entre documentos, autores e instituciones. La metodología empleada en este estudio, actualmente replicada en cinco universidades, ha demostrado ser portable y adaptable a distintos contextos institucionales, lo que permite ampliar el análisis hacia nuevas dimensiones, como el estudio de citas, afiliaciones e índices bibliométricos más verificables. Esto resulta fundamental en el campo de la Documentación, donde trazabilidad, interoperabilidad y precisión de los registros son esenciales para asegurar una gestión del conocimiento sostenible, reutilizable y de calidad en la información consultada.

## REFERENCIAS BIBLIOGRÁFICAS

- Agencia Nacional de Evaluación de la Calidad y Acreditación. (2024). *Resolución de 20 de marzo de 2024, por la que se aprueban los criterios de evaluación y requisitos mínimos de referencia de los méritos y competencias requeridos para obtener la acreditación a Catedrática o Catedrático de Universidad y a Profesora o Profesor Titular de Universidad*. <https://participa.aneca.es/processes/evaluacion-meritos-competencias/f/3/ANECA+7ANECA+7partici>
- Choi, A. J. y Xin, X., (2021). Data Curation in Practice: Extract Tabular Data from PDF Files Using a Data Analytics Tool. *Journal of eScience Librarianship*, 10(3), 10. <https://doi.org/10.7191/jeslib.2021.1209>
- Fundación Española para la Ciencia y la Tecnología (FECYT). (2022). *Publons se integra en Web of Science*. Recuperado el 3 de mayo de 2025, de <https://www.recursoscientificos.fecyt.es/noticias/publons-se-integra-en-web-science>
- Fundación Española para la Ciencia y la Tecnología. (2022). *FECYT renueva el Sello de la Calidad Editorial y Científica a 514 revistas españolas*. Recuperado el 3 de mayo de 2025 de <https://www.fecyt.es/actualidad/fecyt-renueva-el-sello-de-la-calidad-editorial-y-cientifica-514-revistas-espanolas>
- Fundación Española para la Ciencia y la Tecnología. (2025). *Guía de evaluación: Novena edición de evaluación de la calidad editorial y científica de las revistas académicas españolas*. [https://evaluacionarce.fecyt.es/Publico/Bases/\\_\\_\\_Recursos/2025\\_9\\_convocatoria\\_GuiaEvaluacion.pdf](https://evaluacionarce.fecyt.es/Publico/Bases/___Recursos/2025_9_convocatoria_GuiaEvaluacion.pdf)
- García García-Doncel, J. (2023, agosto 7). Curado de datos. Data Science. Recuperado el 13 de febrero de 2025 de <https://dat-science.com/curado-de-datos/>
- ISSN International Standard Serial Number. (s. f.). Consultar la tabla de números ISSN-L. Recuperado el 30 de mayo de 2025, de <https://www.issn.org/es/servicios-y-prestaciones/servicios-en-linea/consultar-la-tabla-de-numeros-issn-l/>
- Khder, M. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and Its Applications*, 13, 145-168. <https://doi.org/10.15849/IJASCA.211128.11>
- Marín-Arraiza, P. (2022). Madurez de sistemas de identificadores persistentes: Oportunidades en el contexto español. *Anuario ThinkEPI*, 16. <https://doi.org/10.3145/thinkepi.2022.e16a06>

- Minamiyama, Y., Takeda, H., Hayashi, M., Asaoka, M. y Yamaji, K. (2024). A study on formalizing the knowledge of data curation activities across different fields. *PLOS ONE*, 19(4), e0301772. <https://doi.org/10.1371/journal.pone.0301772>PLOS
- Parmiggiani, E., Amagyei, Nana Kwame y Kollerud, S. K. S. (2024). Data curation as anticipatory generification in data infrastructure. *European Journal of Information Systems*, 33(5), 748-767. <https://doi.org/10.1080/0960085X.2023.2232333>
- Peng, G., Downs, R. R. y Wyborn, L. (2022). Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets. *Data Science Journal*, 21(1), 8. <https://doi.org/10.5334/dsj-2022-008>
- Rose, M. E. y Kitchin, J. (2024). pybliometrics: Python-based API-Wrapper to access Scopus. pybliometrics. <https://pybliometrics.readthedocs.io/en/stable/>
- SciELO. (2023). *Guía de curación de datos de investigación para equipos editoriales*. Scielo. <https://www.scielo.org/es/sobre-el-scielo/scielo-data-es/>
- Torres-Salinas, D. y Arroyo-Machado, W. (2022). APIs en contextos bibliométricos: Introducción básica y corpus exhaustivo. *Anuario ThinkEPI*, 16. <https://doi.org/10.3145/thinkepi.2022.e16a09>
- Vélez-Estévez, A., Pérez, I. J., García-Sánchez, P., Moral-Munoz, J. A. y Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, 60(4), 103385. <https://doi.org/10.1016/j.ipm.2023.103385>
- Yakel, E. (2007). Digital curation. *OCLC Systems and Services: International digital library perspectives*, 23(4), 335-340.

## CONVERGÊNCIAS ENTRE A FOLKSONOMIA E A CURADORIA SOCIAL E SUAS CONTRIBUIÇÕES PARA A CIÊNCIA CIDADÃ


### *CONVERGENCIAS ENTRE LA FOLKSONOMÍA Y LA CURADURÍA SOCIAL Y SUS CONTRIBUCIONES A LA CIENCIA CIUDADANA*

### *CONVERGENCES BETWEEN FOLKSONOMY AND SOCIAL CURATION AND THEIR CONTRIBUTIONS TO CITIZEN SCIENCE*

Gabriela DE OLIVEIRA SOUZA<sup>1</sup> y Maria José VICENTINI JORENTE<sup>2</sup>

<sup>1</sup> *Universidade Estadual Paulista (UNESP). Brasil*

[gabriela.oliveira@unesp.br](mailto:gabriela.oliveira@unesp.br)

 <https://orcid.org/0000-0001-7519-6624>

<sup>2</sup> *Universidade Estadual Paulista (UNESP). Brasil*

[mj.jorente@unesp.br](mailto:mj.jorente@unesp.br)

 <https://orcid.org/0000-0002-0492-0918>

RESUMO: Práticas e projetos baseados na Ciência Cidadã permitem a participação de sujeitos não especialistas em atividades de pesquisa científica, que geram grandes volumes de dados e informações. Para apoiar a Ciência Cidadã no que diz respeito ao acesso, compartilhamento e preservação da informação, são necessárias as técnicas e metodologias provenientes da Curadoria Digital e, mais especificamente, é necessária a convergência de uma de suas facetas colaborativas – a Curadoria Social –, que envolve a participação dos sujeitos informacionais ao longo de todo o processo curatorial. Processos como a Folksonomia, uma forma de categorização ou indexação colaborativa em ambientes Web, quando convergidos à Curadoria Social, podem apresentar novas possibilidades para os profissionais da informação e para a Ciência Cidadã de modo geral. Tal convergência faz emergirem práticas colaborativas que contribuem diretamente com a Ciência Cidadã, como o *Crowdsourcing*. O objetivo do presente trabalho é analisar como a Folksonomia, convergida à Curadoria Social, pode contribuir com a Ciência Cidadã. A metodologia aplicada no presente estudo é de natureza qualitativa e de tipologia descritiva e exploratória, com base em uma revisão de literatura sobre as temáticas propostas. Após o levantamento bibliográfico, foi realizada uma revisão de literatura e seleção da bibliografia significativa para o estudo, com o intuito de construir um referencial teórico significativo sobre o tema. O número de artigos recuperados demonstra a necessidade de estudos que envolvam a

Curadoria Social e práticas colaborativas como a Folksonomia no contexto da Ciência Cidadã. A Folksonomia, se convergida às curadorias digital e social, pode aproximar as comunidades de interesse da Ciência Cidadã, de modo que uma abordagem social da Curadoria Digital pode promover a Ciência Cidadã, uma vez que a Curadoria Social torna os sujeitos informacionais curadores de conteúdos, e não apenas consumidores.

PALAVRAS-CHAVES: Ciência Cidadã; curadoria digital; curadoria social; folksonomia; *crowdsourcing*; informação e tecnologia.

RESUMEN: Las prácticas y los proyectos basados en la ciencia ciudadana permiten la participación de personas no especializadas en actividades de investigación científica que generan grandes volúmenes de datos e información. Para apoyar la ciencia ciudadana en términos de acceso, intercambio y preservación de información, son necesarias técnicas y metodologías de curación digital y, más específicamente, es necesaria la convergencia de una de sus facetas colaborativas: la curación social, que implica la participación de los sujetos de información a lo largo de todo el proceso curatorial. Procesos como la folksonomía, una forma de categorización o indexación colaborativa en entornos web, al converger con la curación social, pueden presentar nuevas posibilidades para los profesionales de la información y para la ciencia ciudadana en general. Dicha convergencia da lugar a prácticas colaborativas que contribuyen directamente a la ciencia ciudadana, como el *Crowdsourcing*. El objetivo de este trabajo es analizar cómo la folksonomía, convergente con la curación social, puede contribuir a la ciencia ciudadana. La metodología aplicada en este estudio es de naturaleza cualitativa y descriptiva y exploratoria, basada en una revisión bibliográfica sobre los temas propuestos. Tras la revisión bibliográfica, se realizó una selección bibliográfica y se seleccionó la bibliografía relevante para el estudio, con el objetivo de construir un marco teórico significativo sobre el tema. El número de artículos recuperados demuestra la necesidad de estudios que involucren la curación social y prácticas colaborativas como la folksonomía en el contexto de la ciencia ciudadana. La folksonomía, al converger con la curación digital y social, puede acercar a las comunidades de interés a la ciencia ciudadana, de modo que un enfoque social de la curación digital pueda promover la ciencia ciudadana, ya que la curación social convierte a los sujetos de información en curadores de contenido y no solo en consumidores.

PALABRAS CLAVE: ciencia ciudadana; curación digital; curación social; folksonomía; *crowdsourcing*; tecnologías de la información.

ABSTRACT: Practices and projects based on Citizen Science allow the participation of non-specialists in scientific research activities that generate large volumes of data and information. To support Citizen Science in terms of access, sharing and preservation of information, techniques and methodologies from Digital Curation are necessary and, more specifically, the convergence of one of its collaborative facets is necessary – Social Curation –, which involves the participation of information subjects throughout the entire curatorial process. Processes such as Folksonomy, a form of collaborative categorization or indexing in Web environments, when converged with Social Curation, can present new possibilities for information professionals and for Citizen Science in general. Such convergence gives rise to collaborative practices that directly contribute to Citizen Science, such as Crowdsourcing. The objective of this paper is to analyze how Folksonomy, converged with Social Curation, can contribute to Citizen Science.

The methodology applied in this study is qualitative in nature and descriptive and exploratory in nature, based on a literature review on the proposed themes. After the bibliographic survey, a literature review and selection of the relevant bibliography for the study were carried out, with the aim of constructing a meaningful theoretical framework on the topic. The number of articles retrieved demonstrates the need for studies involving Social Curation and collaborative practices such as Folksonomy in the context of Citizen Science. Folksonomy, if converged with digital and social curation, can bring communities of interest closer to Citizen Science, so that a social approach to Digital Curation can promote Citizen Science, since Social Curation makes information subjects curators of content, and not just consumers.

KEYWORDS: Citizen Science; Digital Curation; Social Curation; Folksonomy; Crowdsourcing; Information and Technology.

## 1. INTRODUÇÃO

As transformações das Tecnologias de Informação e Comunicação (TIC) ao longo do tempo proporcionaram mudanças significativas na maneira com que os indivíduos trabalham e solucionam problemas. No entanto, é necessário que se faça um melhor uso das tecnologias para aumentar as competências humanas em solucionar problemas difíceis e complexos. As soluções para tais problemas podem ser encontradas nos conhecimentos e competências coletivas, que com o auxílio das tecnologias, ampliam as capacidades humanas em coletar informações, manipulá-las, compartilhá-las e, a partir de tal processo, gerar conhecimentos e colocá-los em prática.

É nesse sentido que a Ciência Cidadã se faz necessária na contemporaneidade, pois envolve a participação das comunidades de interesse em atividades científicas (Haklay *et al.*, 2021). Práticas e projetos que envolvem a Ciência Cidadã permitem e incentivam a participação de sujeitos não especialistas em atividades de pesquisa científica, que geralmente exigem grandes quantidades de dados e informações, e são facilitadas ou mediadas pelas tecnologias (Hidalgo *et al.*, 2021).

Na Ciência da Informação, tais práticas se fundamentam no paradigma pós-custodial que, segundo Silva e Ribeiro (2020), abrange as novas formas de tratamento da informação, além de novos modelos científicos que contemplam a complexidade do contexto atual. Assim, destaca-se a ampliação dos espaços de atuação dos profissionais da informação, que não atuam somente em algumas funções fixas, mas possuem um amplo leque de possibilidades de atuação devido à complexidade do novo paradigma.

Além disso, o paradigma pós-custodial busca se ajustar às novas demandas informacionais contemporâneas, e reconhece a informação enquanto um elemento humano, social e dinâmico, disponibilizado em um suporte. O principal foco do paradigma é o acesso à informação, pois considera que apenas o acesso à informação legítima e justifica sua custódia e preservação (Silva e Ribeiro, 2020).

As atividades colaborativas geram grandes volumes de dados e informações, de modo que para apoiar a Ciência Cidadã no que diz respeito ao acesso, compartilhamento e preservação

da informação, são necessárias as técnicas e metodologias provenientes da Curadoria Digital (CD), que compreende a gestão, a preservação e o acesso aos objetos informacionais digitais ao longo de todo seu ciclo de vida. Mais especificamente, é necessária a convergência de uma de suas facetas colaborativas – a Curadoria Social (CS) –, que envolve a participação dos sujeitos informacionais e das comunidades de interesse ao longo de todo o processo curatorial.

Além disso, processos como a Folksonomia, uma forma de categorização ou indexação colaborativa em ambientes Web, quando convergidos à Curadoria Social (CS) podem apresentar novas possibilidades para os profissionais da informação e para a ciência cidadã de modo geral. Tal convergência faz emergirem práticas colaborativas que contribuem diretamente com a Ciência Cidadã, como o *Crowdsourcing*, por exemplo.

Nesse contexto, o objetivo do presente trabalho é analisar como a Folksonomia, convergida à Curadoria Social, pode contribuir com a Ciência Cidadã.

## 2. Metodologia

A metodologia aplicada no presente estudo é de natureza qualitativa e de tipologia descritiva e exploratória, com base em uma revisão de literatura sobre as temáticas propostas.

A pesquisa qualitativa utiliza diferentes abordagens para compreender diversos fenômenos sociais, na qual o pesquisador torna-se, concomitantemente, sujeito e objeto das pesquisas. Um estudo de natureza qualitativa preocupa-se com elementos da realidade que não podem ser quantificados, e busca compreender e explicar a dinâmica das relações sociais (Gerhardt e Silveira, 2009).

Quanto à sua tipologia, este estudo pode ser considerado como descritivo e exploratório. A tipologia descritiva busca descrever as características de um fenômeno específico, as relações entre variáveis, ou ainda a natureza de tais relações (Gil, 2008). Há estudos que, apesar de suas características descritivas, buscam, também, apresentar novas perspectivas acerca de seu objeto de estudo ou soluções práticas para seus problemas de pesquisa, de modo que também podem ser classificadas como exploratórios (Gil, 2008), como o estudo em questão.

A pesquisa exploratória busca coletar informações sobre o objeto de pesquisa e tem como principal objetivo o desenvolvimento e esclarecimento de conceitos e ideias para a construção de hipóteses e problemas mais precisos (Severino, 2007; Gil, 2008). Grande parte das pesquisas exploratórias utilizam como principais métodos e procedimentos o levantamento bibliográfico e documental, as entrevistas e os estudos de caso; e apresentam inicialmente uma investigação mais ampla, que no decorrer da pesquisa se delimita e faz emergirem novas discussões (Gil, 2008).

Nesse contexto, foi realizado um levantamento bibliográfico nas bases de dados Brapci, Scielo e Scopus, devido à sua relevância para a Ciência da Informação no âmbito nacional e internacional. A busca nas bases de dados retornou um número significativo de artigos, no entanto, os materiais foram selecionados para a revisão de literatura de acordo com o protocolo apresentado no quadro a seguir (Quadro 1).

Quadro 1. Protocolo para revisão de literatura

<b>Fontes de informação</b>	Bases de dados Brapci, SciElo e Scopus Material publicado por Thomas Vander Wal e pelo <i>Digital Curation Centre</i> (DCC)
<b>Expressões de busca</b>	«Digital Curation» AND «Social Curation» «Digital Curation» AND «Social Curation» AND «Citizen Science» «Citizen Science» AND «Information Science» Folksonomy AND Crowdsourcing Crowdsourcing AND «Citizen Science»
<b>Recorte temporal</b>	2014-2025
<b>Idioma</b>	Português, inglês e espanhol
<b>Tipo de material</b>	Artigos e capítulos de livros
<b>Acesso</b>	Acesso aberto
<b>Critérios para seleção</b>	Leitura do título, resumo e palavras-chave para identificação da relevância para o estudo.

*Nota:* Elaborado pelas autoras.

O Quadro 1 apresenta o protocolo utilizado para a seleção da literatura. Destaca-se que além dos artigos e capítulos recuperados nas bases de dados, também foram considerados os materiais publicados por Thomas Vander Wal, criador do termo Folksonomia, e as publicações do *Digital Curation Centre* (DCC), referência em estudos sobre Curadoria Digital. As expressões de busca foram elaboradas em língua inglesa para ampliar o escopo da busca em âmbito internacional, e o recorte temporal foi delimitado entre os anos de 2014 e 2025. Além disso, a busca se restringiu a artigos e capítulos de livro em acesso aberto. Para a seleção da bibliografia, foram analisados o título, o resumo e as palavras-chave dos textos, para identificar sua relevância para o presente estudo. Desse modo, o quadro a seguir (Quadro 2), apresenta o número de artigos recuperados nas bases de dados em cada expressão de busca.

Quadro 2. Resultados da busca nas bases de dados

Expressão de busca	Bibliografia recuperada nas bases de dados		
	Brapci	SciElo	Scopus
«Digital Curation» AND «Social Curation»	27	7	10
«Digital Curation» AND «Social Curation» AND «Citizen Science»	1	0	1
«Citizen Science» AND «Information Science»	70	54	94
Folksonomy AND Crowdsourcing	1	1	14
Crowdsourcing AND «Citizen Science»	1	1	356

*Nota:* Elaborado pelas autoras.

O número de artigos recuperados demonstra a necessidade de estudos que envolvam a Curadoria Social e práticas colaborativas como a Folksonomia no contexto da Ciência Cidadã.

Após o levantamento bibliográfico, foi realizada uma revisão de literatura e seleção da bibliografia significativa para o estudo, com o intuito de construir um referencial teórico significativo sobre o tema.

### 3. CURADORIA DIGITAL E CURADORIA SOCIAL NA CIÊNCIA CIDADÃ

A Ciência Cidadã abrange a participação das comunidades de interesse de modo geral em atividades de pesquisa científica. Embora o termo tenha sido cunhado na década de 1990 e se popularizado desde então, as práticas da Ciência Cidadã são anteriores a essa data, e contam com a colaboração de sujeitos especialistas e não especialistas para construir novos conhecimentos que beneficiem não somente os cientistas e as ciências a nível teórico, mas, principalmente, a sociedade como um todo de forma aplicada. O potencial da Ciência Cidadã está nos efeitos sobre a sociedade de modo geral, ao responder questões científicas, gerar dados válidos e, ainda, impulsionar a mudança social por meio da participação e da colaboração (Vohland *et al.*, 2021).

A consolidação das Ciência Cidadã em diversas áreas do conhecimento faz emergir a necessidade de diretrizes, padrões e vocabulários que sejam capazes de abranger a complexidade das práticas colaborativas. Além disso, a heterogeneidade é uma das características que enriquece as atividades e projetos da Ciência Cidadã, mas oferece desafios e demandas específicas, como recursos, plataformas e aplicativos que proporcionem uma melhor comunicação e interação entre especialistas e não especialistas. A comunicação é essencial no contexto da Ciência Cidadã, uma vez que, se a comunicação com os participantes não for realizada de maneira eficiente, eficaz e efetiva, há grande probabilidade que o projeto não alcance os objetivos esperados (Vohland *et al.*, 2021).

Devido ao grande volume de dados e informações gerados em projetos e práticas da Ciência Cidadã, o uso das Tecnologias de Informação e Comunicação (TIC) torna-se essencial para proporcionar o acesso, a preservação e o compartilhamento do conhecimento produzido colaborativamente.

O crescimento exponencial da produção de objetos informacionais digitais – tanto nativos quanto digitalizados –, acarretou a necessidade de novas práticas e metodologias para seu tratamento, preservação, acesso e compartilhamento a longo prazo. No entanto, apenas estratégias e práticas de preservação não seriam suficientes para proporcionar o acesso, o compartilhamento e a salvaguarda das informações contidas em tais objetos ao longo de todo seu ciclo de vida (Silva e Siebra, 2019).

Diante da complexidade no tratamento das informações disponíveis no cenário contemporâneo dígito-virtual, a Curadoria Digital (CD) emerge enquanto uma área que apresenta um conjunto de ações contínuas e metodologias com o intuito de otimizar o acesso, a preservação e o compartilhamento de informações (Jorente, Silva e Padua, 2021). De acordo com o *Digital*

*Curation Centre* (DCC), a CD compreende a gestão e a preservação de objetos informacionais digitais a longo prazo, por meio de ações que buscam manter, preservar e agregar valor aos dados digitais ao longo de todo seu ciclo de vida (DCC, 2004).

Segundo Gillian Oliver e Ross Harvey (2016), os objetos digitais possuem especificidades que exigem sua gestão ativa, de modo que diversos fatores ocasionam a necessidade da CD, conforme apresentado no quadro a seguir (Quadro 3).

Quadro 3. Fatores que implicam na necessidade de Curadoria Digital

<b>Obsolescência tecnológica</b>	Atualização frequente de hardware e software, que pode impossibilitar o acesso à informação em sistemas e formatos desatualizados.
<b>Fragilidade tecnológica</b>	Objetos informacionais digitais podem se tornar inacessíveis caso uma parte de seus dados seja alterada ou corrompida.
<b>Falta de compreensão de boas práticas de preservação e acesso</b>	Desconhecimento de estratégias e metodologias para a preservação e acesso a longo prazo, bem como dos elementos necessários para tratamento e descrição dos objetos digitais, como padrões de metadados e modelos conceituais para o planejamento da preservação.
<b>Falta de recursos ou recursos inadequados</b>	Unidades de informação podem não dispor de recursos para as atividades de preservação e para a capacitação dos profissionais.
<b>Falta de infraestrutura</b>	Incertezas sobre a infraestrutura técnica e organizacional para o tratamento, salvaguarda e compartilhamento dos objetos digitais.
<b>Grande volume de dados e informações</b>	Produção constante de grandes conjuntos de dados que necessitam de tratamento e armazenamento.

Nota: Elaborado pelas autoras com base em Oliver e Harvey (2016).

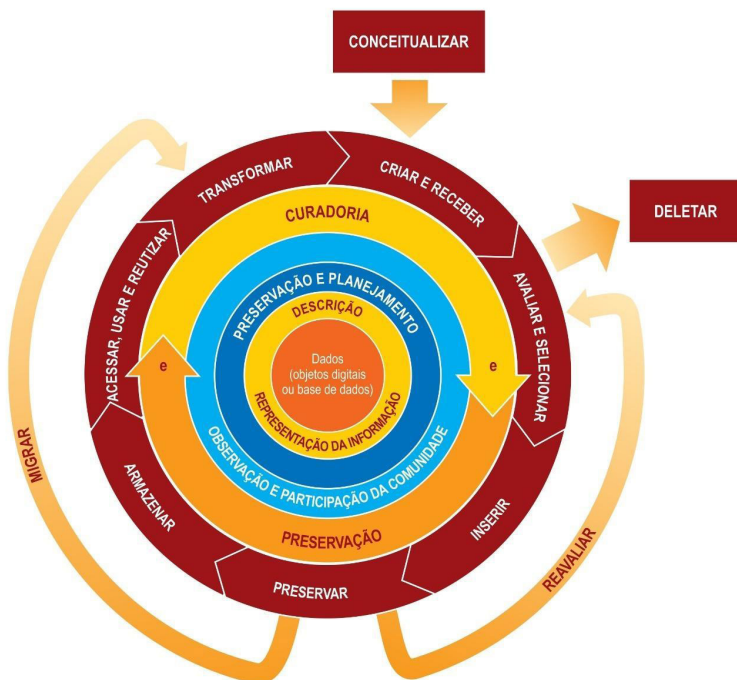
Conforme o quadro anterior, a complexidade do atual contexto de produção, tratamento, preservação, acesso e compartilhamento de informações faz emergir a necessidade da CD que, por sua vez, pode ajudar a solucionar tais impasses. A CD compreende ações e práticas contínuas que abrangem todo o ciclo de vida dos objetos informacionais digitais, desde sua criação, até seu armazenamento, acesso e possível descarte, o que possibilita o planejamento de todo o processo. Desse modo, é possível planejar a migração de hardware e software, a avaliação e limpeza de dados, a inserção de metadados descritivos e de preservação, a transferência para repositórios confiáveis e as formas de acesso. Também é possível prever quais recursos técnicos e tecnológicos serão necessários para a execução do processo de CD de forma eficiente, eficaz e efetiva.

A CD se desenvolveu a partir de conceitos da museologia referentes à curadoria de museus de maneira geral e à processos de preservação digital na curadoria de dados, e os convergiu em ações relacionadas à criação, preservação, compartilhamento e acesso à informação na Web (Santos, 2014). Assim, para ser desenvolvida de forma eficiente e eficaz, a CD deve ser baseada em modelos conceituais (Oliver e Harvey, 2016).

Sarah Higgins (2008) apresentou o modelo do Ciclo de Vida da Curadoria Digital do DCC, uma representação gráfica das etapas e ações da CD, compreendida como um processo

contínuo e cíclico. De modo geral, o modelo é indicativo e não exaustivo, e pode ser aplicado em diferentes instituições (Higgins, 2008). As ações realizadas ou não em cada etapa do Ciclo de Vida da Curadoria Digital influenciam diretamente na eficácia de todo o processo, de modo que o modelo também pode ser utilizado como uma lista de verificação, para auxiliar na gestão do processo de CD (Oliver e Harvey, 2016). A figura a seguir (Figura 1) apresenta uma versão traduzida do modelo em questão.

Figura 1. Modelo do Ciclo de Vida da Curadoria Digital



Nota: Traduzido de Higgins (2008).

O Modelo do Ciclo de Vida da Curadoria Digital (Figura 1) é dividido em três tipos de ações: ações essenciais, ações sequenciais e ações ocasionais (Higgins, 2008). As ações essenciais, representadas nas cores laranja, amarelo e azul na figura, abrangem: dados; descrição e representação da informação; preservação e planejamento; participação e observação da comunidade; e curadoria e preservação. As ações sequenciais, representadas em vermelho na figura, compreendem: conceitualizar; criar e receber; avaliar e selecionar; inserir; preservar; armazenar; acessar, usar e reutilizar; e transformar. Já as ações ocasionais, representadas por meio das setas em tons alaranjados, são: deletar, reavaliar e migrar.

No contexto da Ciência Cidadã, destaca-se a ação *participação e observação da comunidade*, do ciclo de vida da CD que, de acordo com Higgins (2008), corresponde ao desenvolvimento colaborativo de padrões, recursos e softwares. No entanto, ainda que a participação da comunidade faça parte de uma ação, o termo «comunidade» é frequentemente relacionado a uma comunidade especializada, ou seja, profissionais da informação ou equipes técnicas de tecnologia da informação (Brayner, 2018). Esta definição de comunidade pode não contemplar os sujeitos informacionais de modo geral, ou seja, os sujeitos não especialistas. Contudo, para este trabalho, considera-se como participação da comunidade a atuação de qualquer sujeito informacional nas ações de CD, pois a colaboração dos internautas pode apresentar diversos benefícios para as unidades de informação.

De acordo com Siebra (2019), ainda que existam diversos modelos conceituais para a aplicação da CD, como o Ciclo de Vida da Curadoria Digital do DCC, devido à diversidade e volume de objetos informacionais digitais criados na contemporaneidade, não há uma única forma de se aplicar a CD. Nesse sentido, deve-se considerar as particularidades dos objetos dígito-virtuais a serem tratados, bem como da instituição e da comunidade de interesse a qual os objetos informacionais se destinam (Siebra, 2019).

A colaboração das comunidades envolvidas no processo da CD – criadores de dados, profissionais, internautas, e qualquer comunidade envolvida direta ou indiretamente – é uma das chaves para o seu desenvolvimento de forma eficiente e eficaz. Isto porque a aplicação da CD apresenta diversos desafios – institucionais, estruturais, tecnológicos, entre outros –, e a contribuição das comunidades de interesse pode fornecer respostas úteis para a solução de problemas (Oliver e Harvey, 2016).

Enquanto uma área complexa, a Curadoria Digital possui diversas facetas, uma delas é a Curadoria Social (CS), que pode solucionar a necessidade de participação apresentada por Oliver e Harvey (2016). A principal característica que a diferencia da CD é seu propósito social, ou seja, a Curadoria Social sempre é realizada tendo em vista o aprendizado, a colaboração e a participação das comunidades de interesse (Seitzinger, 2014).

A CS envolve a descoberta, seleção, *harvesting* (coleta) e compartilhamento de objetos informacionais digitais. A descoberta implica na configuração, monitoramento e ajuste dos fluxos informacionais para a descoberta dos objetos informacionais digitais a serem compartilhados. A seleção contempla o processo de tomada de decisão dos curadores ao selecionar os objetos informacionais que devem ser compartilhados e quais devem ser apenas preservados. O *harvesting* – ou coleta – consiste na criação de coleções, inserção de etiquetas (Folksonomia), comentários e outras formas de identificação dos objetos, bem como a coleta do resultado da atividade colaborativa. Por fim, no compartilhamento, os curadores compartilham e disponibilizam as coleções e objetos informacionais digitais em plataformas sociais, para que sejam descobertos por outros internautas (Seitzinger, 2014).

A CS também busca agregar significado às coleções dígito-virtuais por meio da interação das comunidades de interesse em ambientes dígito-virtuais colaborativos (Sabharwal, 2015). Segundo Miranda, Arruda, Viana e Santos (2024) a CS compreende um processo de curadoria de conteúdos que não depende apenas dos profissionais da informação, mas

principalmente da atuação dos sujeitos informacionais, vistos, desse modo, como curadores de conteúdos «que buscam compartilhar o conhecimento de forma coletiva, colaborando para a criação de uma inteligência coletiva acerca de uma temática ou abordagem» (Miranda, Arruda, Viana e Santos, 2024, p. 188).

Além disso, a CS propõe uma abordagem participativa da CD, e busca envolver as comunidades de interesse e a sociedade como um todo nos processos curatoriais. Tal abordagem pode ocorrer por meio da abertura de acervos de bibliotecas, arquivos e museus em ambientes dígito-virtuais nos quais a comunidade pode colaborar por meio do compartilhamento de informações e da interação com os acervos, o que torna horizontais os processos de Curadoria Digital (Batista, Jorente e Padua, 2022).

A relação entre preservação e acesso na CD pode ocorrer de duas formas: organizar e preservar o conhecimento com o intuito de proporcionar o acesso, ou primeiramente proporcionar o acesso a um conjunto de informações sem tratamento e depois preservá-las por meio do resultado de atividades colaborativas. A CS representa o segundo processo, ou seja, a inversão dos processos tradicionais da CD em uma perspectiva que parte do acesso para a preservação - inversão que segue a perspectiva do paradigma pós-custodial da Ciência da Informação, conforme apresentado por Silva e Ribeiro (2020), que pressupõe que apenas o acesso justifica e legitima a preservação. Também se enquadra no último paradigma proposto por Silva e Ribeiro, em seu texto de 2020, o paradigma político-ideológico e sociocultural, no qual a cultura e os fluxos informacionais são trabalhados a partir das necessidades sociais dos sujeitos informacionais.

#### 4. CONVERGÊNCIAS ENTRE A FOLKSONOMIA E A CURADORIA SOCIAL

No contexto da Curadoria Social destacam-se atividades e práticas colaborativas, como a Folksonomia, termo criado para designar o resultado da etiquetagem pessoal de informações e objetos digitais para sua própria recuperação. A Folksonomia apresenta três elementos principais: a etiqueta, o objeto marcado por uma etiqueta, e a identidade, princípios fundamentais para a compreensão do objeto etiquetado e do processo como um todo. Além de categorizar ou classificar um objeto digital, os indivíduos estabelecem conexões entre diversos objetos e fornecem novos significados a eles, de modo que o valor da etiquetagem se refere ao uso do vocabulário próprio dos sujeitos informacionais, que imprimem ao objeto significado proveniente de sua própria compreensão (Wal, 2007), ou seja, uma semântica, o que viabiliza que a representação e a apresentação da informação sejam realizadas de forma eficiente, eficaz e efetiva.

A Folksonomia é uma atividade social, uma vez que os internautas classificam um objeto informacional e utilizam as etiquetas já existentes, para que outras pessoas também possam encontrá-lo (Wal, 2005). Além disso, se popularizou em resposta à crescente interação dos

internautas em ambientes dígito-virtuais, e reflete o vocabulário das comunidades de interesse, por ser uma forma de indexação simples em linguagem natural que agrega valor à navegação em ambientes web (Yu e Chen, 2020).

Como qualquer processo, a Folksonomia apresenta vantagens e desvantagens. Entre as vantagens, destacam-se a exaustividade na indexação dos objetos informacionais digitais, a facilidade na recuperação da informação e a participação da comunidade de interesse no processamento técnico da informação. As desvantagens seriam a ausência no controle do vocabulário e a repetição de termos (Sundström e Moraes, 2019). No entanto, ainda que existam alguns questionamentos acerca da validade de uma indexação realizada de forma livre pelos internautas, diferentes experiências com tal prática em ambientes Web demonstram que a Folksonomia pode contribuir de forma significativa com a qualidade dos metadados de objetos informacionais digitais e, conseqüentemente, com a recuperação da informação (Yu e Chen, 2020).

Autores como Golub, Lykke e Tudhope (2014), por exemplo, destacaram a Folksonomia enquanto resultado de uma marcação social realizada em ambientes Web que apresenta como principais vantagens: o baixo custo, pois é uma forma de indexação realizada gratuitamente pelos próprios internautas; a facilidade de inserção de novos termos e conceitos, de acordo com a colaboração das comunidades de interesse; e a garantia de uso, uma vez que os internautas inserem as etiquetas que consideram mais apropriadas para suas próprias atividades no sistema ou ambiente dígito-virtual (Golub, Lykke e Tudhope, 2014).

Além disso, a colaboração dos internautas pode enriquecer os instrumentos tradicionais de tratamento da informação como cabeçalhos de assunto, catálogos, tesouros e vocabulários controlados. A Folksonomia não substitui tais instrumentos, nem o trabalho dos profissionais da informação, mas adiciona uma nova faceta aos procedimentos tradicionais, abre novas possibilidades de atuação e proporciona sua aproximação com a comunidade. Isso porque, ainda que os termos controlados formalizem o processamento técnico, o uso das etiquetas produzidas pelas comunidades de interesse pode refletir um consenso semântico sobre determinada área do conhecimento (Arias, Santos e Cervantes, 2023). Desse modo, os aspectos negativos da Folksonomia podem ser solucionados por meio da atuação dos profissionais da informação nos ambientes informacionais digitais, com o intuito de filtrar possíveis ambigüidades, termos inadequados e repetições de termos.

A variação linguística, própria da linguagem natural é, também, uma das vantagens do uso da Folksonomia, uma vez que os vocabulários controlados apresentam um conjunto específico de termos, que podem não fazer parte do vocabulário dos internautas de modo geral. Ao utilizar a Folksonomia em conjunto com vocabulários controlados e tesouros, as etiquetas inseridas pelos internautas podem ser registradas como variantes dos termos controlados e se tornam, conseqüentemente, pontos de acesso que representam o objeto informacional e tornam o processo de recuperação da informação mais dinâmico (Arias, Santos e Cervantes, 2023).

Assim, a aplicação da Folksonomia em serviços de informação pode apresentar uma terminologia mais atualizada sobre diversas áreas do conhecimento. A utilização dos resultados

da Folksonomia em tais terminologias torna-se significativa no contexto da representação e organização da informação, pois o uso da linguagem natural convergida aos instrumentos tradicionais como os tesouros e vocabulários controlados apresenta novas possibilidades de representação e tratamento da informação (Arias, Santos e Cervantes, 2023).

O destaque de todo o processo da Folksonomia são as etiquetas atribuídas aos objetos digitais e o tipo de linguagem utilizada pelos internautas, sujeitos não especialistas que atuam ativamente em ambientes dígito-virtuais colaborativos. Por esse motivo, a Folksonomia também é conhecida como indexação social, marcação social, etiquetagem social, *social bookmarking*, classificação social, representação colaborativa e classificação distribuída (Pinto e Felipe, 2023). Por meio do compartilhamento o internauta constrói novas relações entre os objetos informacionais e os outros sujeitos que também interagem na Web e, desse modo, torna-se ativo e «manifesta a sua subjetividade através do estabelecimento de identidades e percursos informacionais na web» (Assis e Moura, 2013, p. 86).

Nesse cenário, uma das práticas que demonstra a convergência entre a Folksonomia e a Curadoria Social (CS) é o *Crowdsourcing*, caracterizado pelo uso da participação social e da colaboração na solução de problemas. Assim, o *Crowdsourcing* possibilita que as comunidades de interesse manifestem suas competências na solução de problemas complexos, além de incentivar a produção e o compartilhamento de conteúdos, o que o torna uma prática essencialmente social (Chieh, 2019) e que pode ser uma estratégia a ser aplicada pela ciência cidadã.

O termo *Crowdsourcing* foi desenvolvido em 2006 pelo jornalista e pesquisador norte-americano Jeff Howe, para denominar o uso da participação social e da colaboração na solução de problemas e na inovação (Chieh, 2019). O *Crowdsourcing* apresenta práticas que utilizam a inteligência coletiva para produzir inovações, e se caracterizam principalmente pela participação social e colaboração para a solução de problemas em instituições, organizações e empresas públicas ou privadas (Fernandes, 2012). Tais práticas buscam aliar a colaboração de fontes externas ao conhecimento interno de determinada instituição, de modo a reconhecer que as respostas para os problemas podem vir de diferentes lugares e indivíduos (Chieh, 2019).

O *Crowdsourcing* pode ocorrer em diversos ambientes da Web, como plataformas sociais criadas especificamente para esta prática e redes sociais, e pode promover a inclusão social ao oportunizar que as comunidades de interesses manifestem suas competências na solução de problemas, motivando-as a produzir e compartilhar conteúdos (Chieh, 2019).

Fernandes (2012) e Chieh (2019) destacaram as tipologias do *Crowdsourcing*, representadas graficamente na figura a seguir (Figura 2).

Figura 2. Tipologias do Crowdsourcing



Nota: Elaborado pelas autoras com base em Fernandes (2019) e Chieh (2012).

Conforme apresentado no mapa conceitual anterior (Figura 2), os autores apresentaram oito tipologias do *Crowdsourcing* (Chieh, 2012; Fernandes, 2019):

1. *Crowd wisdom*: Tem como base a inteligência coletiva, de modo que tal processo recorre à comunidade de interesse para prever acontecimentos e resultados, ou ainda, encontrar soluções para problemas com maior precisão.
2. *Crowd creation*: A comunidade cria conteúdos de forma colaborativa, ou seja, esta tipologia utiliza a criatividade coletiva.
3. *Crowd voting*: Permite que a comunidade de interesse manifeste suas opiniões por meio da votação ou classificação.
4. *Crowdfunding*: Consiste no financiamento coletivo, ou seja, os internautas podem colaborar financeiramente com diversos projetos ou causas.
5. *Knowledge discovery and management*: Uma comunidade é desafiada a encontrar informações sobre uma temática específica, com o objetivo de ampliar os conhecimentos de uma instituição ou organização sobre a temática proposta.

6. *Broadcast search*: Uma organização ou instituição procura, em meio aos internautas, respostas e soluções para problemas que não foram solucionados internamente, de modo que algumas instituições oferecem prêmios aos internautas que solucionam o problema.
7. *Peer-vetted creative production*: Consiste na abertura da etapa de concepção de um produto ou serviço, com o intuito de encontrar, entre os internautas, ideias que possam auxiliar no processo de criação.
8. *Distributed human intelligence tasking*: Utilizado para processar grandes quantidades de dados ou desenvolver tarefas complexas, que são divididas em partes e realizadas colaborativamente (Chieh, 2012; Fernandes, 2019).

Dessa forma, o *Crowdsourcing* pode ser compreendido enquanto uma prática colaborativa desenvolvida em ambientes dígito-virtuais, com foco na participação das comunidades de interesse cujos membros podem fornecer conhecimentos variados. Caracteriza-se pela heterogeneidade e pelo uso das tecnologias e do conhecimento coletivo para atingir um objetivo em comum (Triques, Santos e Albuquerque, 2023).

Segundo Triques, Santos e Albuquerque (2023), independente da tipologia de *Crowdsourcing* aplicada, estudos acerca de tal prática destacam o potencial da colaboração e do compartilhamento de informações. De acordo com as autoras, este processo de colaboração em ambientes Web desenvolveu-se principalmente a partir da emergência dos ambientes que permitiram a etiquetagem colaborativa, ou seja, a Folksonomia.

Tais práticas colaborativas são essenciais no contexto da Curadoria Social (CS) e da Ciência Cidadã. Nesse sentido, enquanto uma faceta da CD, a CS também pode ser aplicada de forma mais eficiente e eficaz se baseada em modelos conceituais. No entanto, o modelo do Ciclo de Vida da Curadoria Digital apresentado pelo DCC apresenta poucas aberturas para atividades colaborativas, de modo que é possível adaptar o modelo para inserir a Folksonomia e a participação da comunidade em geral – o que inclui o *Crowdsourcing* – para que atenda as especificidades da CS e contribua com a Ciência Cidadã.

É possível identificar aberturas para a inserção da Folksonomia nas seguintes ações do ciclo de vida da CD: dados; descrição e representação da informação; participação e observação da comunidade; curadoria e preservação; conceitualizar; criar e receber; acessar, usar e reutilizar; e transformar.

A Folksonomia, enquanto uma forma de indexação colaborativa, gera metadados que identificam os objetos informacionais digitais simples (texto, recursos imagéticos, entre outros) e complexos (bancos de dados e páginas Web). Ao considerar os dados como qualquer tipo de informação em formato digital binário, é possível afirmar que, enquanto processo, a Folksonomia gera dados.

Nas ações de descrição e representação da informação, a Folksonomia pode fornecer as palavras-chave para a representação do conteúdo dos objetos informacionais digitais. Destaca-se a necessidade do uso de padrões de metadados, no entanto, as palavras-chave coletadas na Folksonomia podem ser utilizadas para indexar e representar o conteúdo dos objetos, o que não exclui a relevância do uso dos padrões, mas os complementa.

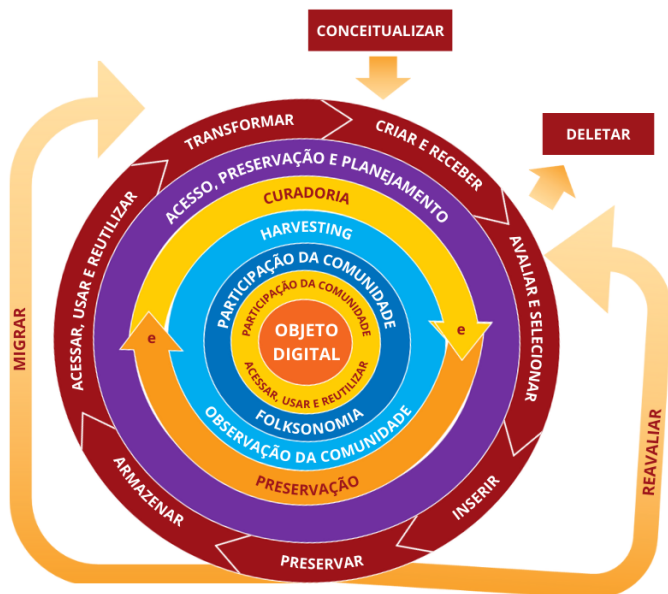
A ação de participação e observação da comunidade, no contexto da Folksonomia, representa a atuação das comunidades de interesse na curadoria dos acervos. Esta ação corresponde à Folksonomia em si, ao processo de etiquetagem colaborativa de objetos informacionais digitais. Considera-se, nesse contexto, que esta ação pode contemplar todas as outras que envolvem a Folksonomia.

Na etapa curar e preservar, a Folksonomia estaria incluída nas ações planejadas para proporcionar a curadoria dos objetos digitais, sendo considerada enquanto um elemento de preservação a ser utilizado em outras ações da CD. Em conceitualizar, a Folksonomia pode ser empregada como um método para a coleta de dados. Já em criar e receber, há o agrupamento dos dados coletados por meio da Folksonomia.

Tradicionalmente, a Folksonomia é encarada como um elemento utilizado para proporcionar o acesso e facilitar a recuperação da informação. Nas ações de acesso, uso e reuso, a Folksonomia assume essa função. A última ação do ciclo de vida da CD em que a Folksonomia pode ser utilizada é a transformação, pois permite a criação de novos objetos informacionais. Por meio dos termos coletados no processo de Folksonomia, os profissionais da informação podem elaborar tesouros e vocabulários controlados para auxiliar nos processos técnicos das unidades de informação.

Nesse contexto, a figura a seguir apresenta a adaptação do Ciclo de Vida da Curadoria Digital, voltado às ações que envolvem a participação da comunidade.

Figura 3. Modelo do Ciclo de Vida da Curadoria Digital voltado à participação da comunidade



Nota: Elaborado pelas com base no DCC Curation Lifecycle Model, de Higgins (2008). Disponível em: <http://hdl.handle.net/11449/239725>

A Figura 3 representa graficamente as ações do Modelo do Ciclo de Vida da Curadoria em que a Folksonomia pode ser aplicada enquanto um elemento de preservação, e foi elaborada com base no modelo original, apresentado por Higgins (2008).

Além das ações anteriormente apontadas, a adaptação do modelo do ciclo de vida da CD apresenta ações colaborativas em outros estágios do ciclo de vida, o que salienta a necessidade da participação dos sujeitos informacionais em grande parte dos processos curatoriais. Ressalta-se que a Folksonomia depende da atuação de indivíduos não especializados, assim, faz-se necessário que o profissional da informação integre essas práticas colaborativas nas unidades de informação.

Assim, foram acrescentadas ou modificadas as seguintes ações: Participação da comunidade – acessar, usar e reutilizar; Participação da comunidade – Folksonomia; *Harvesting* – observação da comunidade; Acesso, preservação e planejamento.

No ciclo adaptado a atividades colaborativas, a participação da comunidade foi incluída em duas ações: participação da comunidade – acessar, usar e reutilizar –; e participação da comunidade – Folksonomia. Na primeira – acessar, usar e reutilizar –, os objetos digitais são compartilhados para que os internautas colaborem no processo de preservação. A princípio, é feita uma descrição simples do objeto digital, depois ele é compartilhado e os internautas podem acessar e utilizar o objeto informacional. Na segunda ação – Folksonomia –, os internautas podem colaborar com o processamento técnico dos objetos informacionais, ao inserir etiquetas a eles.

*Harvesting* – observação da comunidade corresponde à coleta dos termos inseridos pelos internautas por meio da Folksonomia (*harvesting*). Após a coleta, deve ser realizada a análise e a organização dos termos para a construção de instrumentos como vocabulários controlados e tesouros. Também foi acrescentada ao ciclo de vida da CD a ação acesso, preservação e planejamento. Tal ação consiste na aplicação do produto da Folksonomia para o processamento técnico dos objetos digitais, para garantir a preservação. Também compreende, por meio da análise das ações anteriores, o planejamento das próximas ações da CD.

## 5. CONCLUSÕES

A colaboração é um elemento essencial para a ciência cidadã, e para proporcionar a colaboração não apenas de pesquisadores, mas das comunidades de interesse de modo geral, são necessárias estratégias, técnicas e metodologias abertas e participativas, como a Folksonomia e as práticas advindas da Curadoria Social (CS).

A Folksonomia depende da atuação de indivíduos não especializados, de modo que tal participação torna-se essencial diante da necessidade de aproximação entre os sujeitos informacionais e os profissionais da informação. A presença dos sujeitos não especialistas nos processos curatoriais contribui para a recuperação, o compartilhamento e a criação de novos conhecimentos de maneira coletiva e colaborativa, o que contribui também para a promoção da ciência cidadã. A Folksonomia, se convergida às curadorias digital e social, pode aproximar as comunidades de interesse da Ciência Cidadã.

Nesse sentido, uma abordagem social da Curadoria Digital (CD) pode promover a Ciência Cidadã, uma vez que a CS torna os sujeitos informacionais curadores de conteúdos, e não apenas consumidores. Devido à hibridez característica do atual contexto de criação, apresentação, acesso e compartilhamento da informação, novos desafios emergem nos ambientes dígito-virtuais (Jorente, Padua e Nakano, 2019).

A CS, convergida à Folksonomia, pode auxiliar a solucionar parte desses desafios. Considera-se que a Folksonomia e a CS contribuem de forma significativa com a Ciência Cidadã, uma vez que proporcionam a participação dos sujeitos informacionais e das comunidades de interesse. Além disso, o *Crowdsourcing*, enquanto prática colaborativa convergente à Folksonomia, também apresenta diversas possibilidades para a Ciência Cidadã, pois cada uma de suas tipologias pode ser aplicada para uma finalidade distinta.

É importante destacar que a Folksonomia e o *Crowdsourcing* não excluem a atuação do profissional da informação, mas abrem novas possibilidades para sua atuação e o aproxima das comunidades de interesse, sobretudo se consideradas como recursos a ser empregados pela Ciência Cidadã e pela Ciência da Informação na curadoria, acesso e compartilhamento de informações. Desse modo, a perspectiva da Ciência Cidadã e da Curadoria Social poderiam ser mais exploradas pela Ciência da Informação, uma vez que o levantamento realizado em bases de dados retornou um número relativamente pequeno de artigos, quando comparado a outras temáticas, como a própria Curadoria Digital, por exemplo. Por outro lado, isso demonstra que há possibilidades para o desenvolvimento de pesquisas que abordem a temática sob variadas perspectivas, devido ao caráter complexo e interdisciplinar da Ciência da Informação.

Nesse contexto, tanto a Folksonomia quanto o *Crowdsourcing* podem ser importantes recursos nas práticas de CS e, conseqüentemente, da Ciência Cidadã. Experiências de diferentes arquivos, bibliotecas e museus em plataformas colaborativas demonstram que a Folksonomia pode auxiliar no acesso e na recuperação de informação de objetos informacionais digitais que já passaram por um processamento técnico anterior, o que representa a forma tradicional de preservar a informação. No entanto, a inversão do processo, como previsto na CS e na Ciência Cidadã, representa outra forma de preservação e de acesso, advinda dos paradigmas pós-custodial e político-ideológico e sociocultural, que pode aproximar as comunidades de interesse das unidades de informação e das práticas científicas no compartilhamento e na construção do conhecimento coletivo.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Arias, M. I. M., Santos, R. F. dos e Cervantes, B. M. N. (2023). Vocabulário controlado e folksonomia: possíveis aproximações na representação da informação e do conhecimento. *Páginas a&B: Arquivos E Bibliotecas*, 20, 199-212. <https://ojs.letras.up.pt/index.php/paginasueb/article/view/13568>
- Assis, J. H. e Moura, M. A. (2013). Folksonomia: a linguagem das tags. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 18(36), 85-106. <https://brapci.inf.br/index.php/res/v/39062>
- Batista, L. da S., Jorente, M. J. V. e Padua, M. C. (2022). Curación social los procesos de curación digital y la reconstrucción del patrimonio cultural de la humanidad: iniciativas Web 2.0. *Revista EDICIC*, 2(4). <https://doi.org/10.62758/re.v2i4.181>

- Brayner, A. A. (2018). Curadoria digital: novos modelos de participação pública na descrição de conteúdos em instituições culturais. *Revista Ibero-Americana de Ciência da Informação*, 12(1), 53-65. <https://periodicos.unb.br/index.php/RICI/article/view/10521>
- Chieh, L. Y. (2019). *Crowdsourcing: como a sabedoria das multidões pode interessar ao campo de pesquisa e ação da ciência da informação* [Dissertação de Mestrado, Universidade Federal de São Carlos]. <https://repositorio.ufscar.br/items/ca9b3620-7a89-4339-9ec7-980b5c978d13>
- Digital Curation Centre. (2004) *What is digital curation?* <https://www.dcc.ac.uk/about/digital-curation>
- Fernandes, D. B. G. (2012). *O Crowdsourcing: quando? como? e porquê?* [Dissertação de Mestrado, Instituto Universitário de Lisboa]. <https://repositorio.iscte-iul.pt/bitstream/10071/5934/1/Tese%20final.pdf>
- Gerhardt, T. E. e Silveira, D. T. (Org.). (2009) *Métodos de pesquisa*. Editora da Ufrgs.
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Atlas.
- Golub, K., Lykke, M. e Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey Decimal Classification. *Journal of Documentation*, 70(5), 801-828. <https://doi.org/10.1108/jd-05-2013-0056>
- Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S. e Vohland, K. (2021). What Is Citizen Science? The Challenges of Definition. Em K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson e K. Wagenknecht (Eds.), *The Science of Citizen Science* (p. 13). Springer. <https://link.springer.com/book/10.1007/978-3-030-58278-4>
- Hidalgo, E. S., Perelló, J., Becker, F., Bonhoure, I., Legris, M. e Cigarini, A. (2021). Participation and Co-creation in Citizen Science. Em K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson e K. Wagenknecht (Eds.), *The Science of Citizen Science* (p. 199). Springer. <https://link.springer.com/book/10.1007/978-3-030-58278-4>
- Higgins, S. (2008). The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, 3(1) 134-139. [https://www.researchgate.net/publication/220924444\\_The\\_DCC\\_curation\\_lifecycle\\_model](https://www.researchgate.net/publication/220924444_The_DCC_curation_lifecycle_model)
- Jorente, M. J. V., Padua, M. C. e Nakano, N. (2019). O Design da Informação como recurso interdisciplinar da curadoria digital em contextos complexos da Ciência da Informação. *Perspectivas em Ciência da Informação*, 3(24), 35-58. <https://periodicos.ufmg.br/index.php/pci/article/view/22629>.
- Jorente, M. J. V, Silva, S. C. e Padua, M. C. (2021) Digital Curation and Information Design in digital environments: women's museums panorama. *Transinformação*, 33, 1-14. <https://www.scielo.br/j/tinf/a/YFsRxqLrjr6bX3hrQxQxBccGJ/?lang=en#>
- Miranda, M. K. F. de O., Arruda, A. M. A., Viana, A. R. L. e Santos, W. A. de L. (2024). A Curadoria Social e a Competência Crítica em Informação como pressupostos de combate à desinformação: um estudo de caso no YouTube. *Informação & Informação*, 28(2), 180-206. <https://doi.org/10.5433/1981-8920.2023v28n2p180>
- Oliver, G. e Harvey, R. (2016). *Digital Curation*. 2. ed. Ala Neal-Schuman.
- Pinto, P. L. T. da R. e Felipe, C. B. M. (2023). Folksonomia na Netflix: proposta de participação dos usuários na recuperação da informação. *Páginas aEb: Arquivos E Bibliotecas*, 19, 235-256. Obtido de <https://ojs.letras.up.pt/index.php/paginasueb/article/view/13140>
- Sabharwal, A. (2015). *Digital curation in the digital humanities: preserving and promoting archival and special collections*. Elsevier.
- Santos, T. N. C. (2014). *Curadoria digital: o conceito no período de 2000 a 2013* [Dissertação de mestrado, Universidade de Brasília]. Repositório Institucional da UnB. <https://repositorio.unb.br/handle/10482/17324>
- Seitzinger, J. (2015). Curate Me! Exploring online identity through social curation in networked learning. Em *Proceedings of 9 International Conference On Networked Learning* (pp. 412-419). University

- of Edinburgh. [https://www.researchgate.net/publication/282158577\\_Curate\\_Me\\_Exploring\\_Online\\_Identity\\_Through\\_Social\\_Curation\\_in\\_Networked\\_Learning](https://www.researchgate.net/publication/282158577_Curate_Me_Exploring_Online_Identity_Through_Social_Curation_in_Networked_Learning)
- Severino, A. J. (2007). *Metodologia do trabalho científico*. 23. ed. Cortez.
- Siebra, S. de A. (2019). Curadoria Digital: uma área em expansão. *Archeion Online*, 6(2), pp. 1-6. <https://doi.org/10.22478/ufpb.2318-6186.2019v6n2.47089>
- Silva, A. M. e Ribeiro, F. (2020). Ciência da informação trans e interdisciplinar: para a superação de equívocos... Em M. Marques e L. Gomes, *Ciência da informação: visões e tendências*. Imprensa da Universidade de Coimbra. [https://sigarra.up.pt/flup/pt/pub\\_geral.pub\\_view?pi\\_pub\\_base\\_id=414775](https://sigarra.up.pt/flup/pt/pub_geral.pub_view?pi_pub_base_id=414775)
- Silva, F. de M. O. e Siebra, S. de A. (2019). Desmaterializando o documento: contribuições da diplomática para a curadoria digital de documentos arquivísticos digitais. *Archeion Online*, 6(2), pp. 23-44. <https://doi.org/10.22478/ufpb.2318-6186.2019v6n2.46345>
- Triques, M. L., Santos, R. F. dos e Albuquerque, A. C. de. (2023). Aplicações folksonômicas em plataformas colaborativas do patrimônio cultural: análise comparativa dos projetos CrowdHeritage e Arquigrafia. *Palavra Chave*, 12(2), e188. [https://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S1853-99122023000100188](https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1853-99122023000100188)
- Vohland, C., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R. e Wagenknecht, C. (2021). Editorial: The Science of Citizen Science Evolves. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson e K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 1-12). Springer. [https://link.springer.com/chapter/10.1007/978-3-030-58278-4\\_1#citeas](https://link.springer.com/chapter/10.1007/978-3-030-58278-4_1#citeas)
- Wal, T. V. (2005) *Explaining and Showing Broad and Narrow Folksonomies*. <https://www.vanderwal.net/random/entysel.php?blog=1635>
- Wal, T. V. (2007) *Folksonomy*. <http://www.vanderwal.net/folksonomy.html>
- Yu, W. e Chen, J. (2020). Enriching the library subject headings with folksonomy. *The Electronic Library*, 38(2), 297-315.




## LA IMPORTANCIA DE LOS METADATOS EN LA PRESERVACIÓN DE LOS REGISTROS DE ARCHIVO DIGITALES

### *THE IMPORTANCE OF METADATA IN THE PRESERVATION OF DIGITAL ARCHIVE RECORDS*

Pedro Felipy CUNHA DA SILVA<sup>1</sup> y Wagner JUNQUEIRA DE ARAÚJO<sup>2</sup>

<sup>1</sup> *Universidade Federal da Paraíba. Brasil*

[pedrofelipy@hotmail.com](mailto:pedrofelipy@hotmail.com)

 <https://orcid.org/0000-0001-9063-4629>

<sup>2</sup> *Universidade Federal da Paraíba. Brasil*

[wagnerjunqueira.araujo@gmail.com](mailto:wagnerjunqueira.araujo@gmail.com)

 <https://orcid.org/0000-0002-2301-4996>

**RESUMEN:** Este artículo presenta los resultados de una revisión sistemática de la literatura (SLR) sobre registros de archivos digitales, que abarca el período 2020 a 2025. La investigación tuvo como objetivo mapear los enfoques contemporáneos de gestión, preservación, seguridad, uso como evidencia, representación, sistemas y formación profesional relacionados con los registros de archivos digitales. Como actualización de una revisión previa realizada en 2019, se adoptó el protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que cubre las fases de identificación, selección, elegibilidad e inclusión. Se utilizaron las bases de datos Emerald, LISTA (EBSCO), Scopus y Web of Science, con el término de búsqueda «registros digitales» y un marco temporal de cinco años. Luego de aplicar los criterios de elegibilidad y leer los textos en su totalidad, 94 publicaciones conformaron el corpus final de la investigación. Los estudios analizados se organizaron en nueve categorías temáticas: (1) gestión de registros digitales en el sector público; (2) preservación digital; (3) seguridad de los registros digitales; (4) inteligencia artificial aplicada a la gestión de registros; (5) registros digitales como evidencia; (6) capacitación en gestión de registros; (7) sistemas de gestión de registros en el sector público; (8) registros e información digitales; y (9) representación de registros digitales. Los resultados indican la consolidación de nuevos desafíos y perspectivas para la Archivística en la era digital, requiriendo el fortalecimiento de las políticas públicas, la alineación entre la teoría y la práctica profesional, así como la ampliación de la formación técnico-científica de los archivistas frente a las complejidades tecnológicas e informacionales de los registros digitales.

**PALABRAS CLAVE:** archivología; preservación digital; inteligencia artificial; evidencia documental; gestión pública.

**ABSTRACT:** This article presents the findings of a systematic literature review (SLR) on archival digital records, covering the period from 2020 to 2025. The research aimed to map contemporary approaches related to the management, preservation, security, evidentiary use, representation, systems, and professional training regarding digital archival records. As an update of a previous review conducted in 2019, the PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was adopted, encompassing the phases of identification, screening, eligibility, and inclusion. The databases Emerald, LISTA (EBSCO), Scopus, and Web of Science were used, with the search term «digital records» and a five-year time filter. After applying the eligibility criteria and performing full-text analysis, 94 publications comprised the final corpus of the study. The analyzed studies were organized into nine thematic categories: (1) management of digital records in the public sector; (2) digital preservation; (3) digital records security; (4) artificial intelligence applied to records management; (5) digital records as evidence; (6) education and training in records management; (7) records management systems in the public sector; (8) digital records and information; and (9) representation of digital records. The results highlight the emergence of new challenges and perspectives for Archival Science in the digital age, demanding the strengthening of public policies, the alignment between theory and professional practice, and the enhancement of technical and scientific training for archivists in light of the technological and informational complexities of digital records.

**KEYWORDS:** digital records; archival science; digital preservation; artificial intelligence; documentary evidence; public management.

## 1. INTRODUCCIÓN

En las últimas décadas, la creciente digitalización de las actividades humanas ha transformado profundamente la naturaleza, la forma y la función de los registros de archivo. En un escenario marcado por la aceleración tecnológica, la expansión de la información en red y la virtualización de los procesos administrativos, los registros digitales han pasado a ocupar un lugar central en la dinámica organizacional, en los flujos documentales y en las prácticas de gestión pública y privada. Este nuevo contexto impone desafíos conceptuales, técnicos y éticos a la archivística, exigiendo la revisión de los fundamentos teóricos y la reformulación de las prácticas profesionales orientadas a la producción, preservación, seguridad y representación de estos documentos.

Ante este panorama, este artículo presenta los resultados de una revisión sistemática de la literatura sobre registros archivísticos digitales, que abarca el período 2020-2025. Se trata de una actualización de una revisión previa realizada en 2019, con el propósito de identificar tendencias emergentes, mapear enfoques consolidados y discutir los aportes teóricos y prácticos más recientes que se han producido sobre el tema. La obra abarca estudios centrados en gestión, preservación, seguridad, inteligencia artificial, prueba documental, formación profesional, sistemas de información, representación y acceso a registros digitales, componiendo un panorama amplio y multifacético de la producción científica reciente en el área.

El objetivo principal es sistematizar los avances y desafíos que enfrenta la archivística en el tratamiento de los registros digitales, buscando comprender cómo la literatura internacional

ha abordado las complejidades de este objeto informacional. Para tal fin, se utilizó una metodología de revisión sistemática basada en el protocolo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que guio las etapas de identificación, selección y análisis del corpus bibliográfico. Los resultados se organizarán en categorías temáticas que expresen los principales focos de atención y debate presentes en los estudios analizados.

En comparación con la revisión inicial de 2019, en la cual predominaba una preocupación centrada en la estructuración de metadatos y la organización formal de los registros digitales, los estudios más recientes indican un desplazamiento del interés hacia aspectos relacionados con la gestión activa, el uso estratégico y los impactos sociales de estos documentos. Esta evolución temática se evidencia en la creciente presencia de investigaciones que abordan el uso de tecnologías emergentes, como la inteligencia artificial, así como la vinculación de los registros digitales con la transparencia gubernamental, la preservación digital sostenible y la formación profesional crítica en contextos archivísticos complejos.

## 2. METODOLOGÍA

Esta investigación constituye una actualización de la revisión sistemática de la literatura realizada en 2019 sobre registros digitales de archivo, publicada en la *Revista Brasileira de Preservação Digital*. Para garantizar la comparabilidad entre los estudios, se adoptó el mismo protocolo metodológico utilizado en la revisión anterior: PRISMA –Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Este protocolo ofrece una guía consolidada y estructurada para la conducción de revisiones sistemáticas, asegurando rigor, reproducibilidad y transparencia en las fases de identificación, selección, elegibilidad e inclusión de estudios. Su adopción permite un control más estricto de sesgos y una presentación clara del proceso de construcción del corpus documental.

Las búsquedas se realizaron entre los días 1 y 3 de abril de 2025 en las bases de datos Emerald Insight, Scopus, Web of Science y LISTA (EBSCO). Cabe señalar que, debido a la indisponibilidad de la base LISA, utilizada en la revisión de 2019, se optó por incluir la base LISTA como sustituta, dado su enfoque temático convergente y su cobertura en áreas de Bibliotecología y Ciencia de la Información. El término de búsqueda empleado en 2019, «gestión de registros digitales», fue reemplazado por «registros digitales» (*digital records*), a fin de ampliar la sensibilidad de la búsqueda. A todos los resultados se les aplicó un filtro temporal para restringir el corpus al período de 2020 a 2025.

La recuperación de documentos varió según la base consultada, totalizando 158 publicaciones susceptibles de lectura completa. Tras la eliminación de duplicados y documentos inaccesibles, se procedió a la lectura integral de 99 textos. De estos, cinco fueron excluidos por no ajustarse al enfoque temático propuesto, resultando un corpus final compuesto por 94 estudios científicos.

La organización, el almacenamiento y el tratamiento inicial de los datos bibliográficos se realizaron mediante el software Mendeley, utilizando las funcionalidades de importación y gestión de metadatos en formatos RIS y BIBTEX. Posteriormente, se construyó una planilla

de control con campos como autoría, año, país de origen, enfoque temático principal y presencia de tecnologías emergentes, lo que permitió sistematizar y visualizar mejor las dimensiones analíticas de los textos.

La categorización temática de los artículos fue resultado de un proceso de lectura interpretativa y analítica, guiada por los objetivos de la investigación. Se tomaron en cuenta elementos como el enfoque metodológico, los conceptos movilizados, los objetos de estudio y las contribuciones teóricas. Esta clasificación dio origen a nueve categorías analíticas: gestión, preservación, seguridad, inteligencia artificial, evidencia, formación profesional, sistemas de información, representación y acceso a registros digitales. Estas categorías estructuran la presentación de resultados del estudio, permitiendo una visión panorámica y crítica del estado del arte en el área de registros archivísticos digitales en el período investigado.

### 3. RESULTADOS Y DISCUSIÓN

Esta revisión sistemática de la literatura tuvo como objetivo analizar la producción científica sobre registros archivísticos digitales de 2020 a 2025. Para ello, los estudios seleccionados se organizaron en categorías temáticas que expresan los principales ejes de discusión identificados en la literatura. El enfoque adoptado permitió mapear un panorama integral de los aportes teóricos y prácticos que se han desarrollado en el campo de la archivística como respuesta a las transformaciones ocasionadas por la digitalización de documentos.

La categorización de los artículos se realizó con base en una lectura interpretativa de los textos completos, considerando elementos como el alcance temático, los objetos de estudio, los conceptos movilizados y las contribuciones presentadas. Como resultado, se identificaron nueve categorías: (1) Gestión de registros digitales en el sector público; (2) Preservación digital; (3) Seguridad de los registros digitales; (4) Inteligencia artificial aplicada a la gestión de registros; (5) Los registros digitales como prueba; (6) Capacitación en gestión de registros; (7) Sistemas de gestión de registros en el sector público; (8) Registros e información digitales; y (9) Representación de registros digitales.

#### 3.1. GESTIÓN DE REGISTROS DIGITALES EN EL SECTOR PÚBLICO

La producción, el mantenimiento y el uso de registros digitales en contextos gubernamentales implican desafíos técnicos, institucionales y regulatorios. La ausencia de políticas específicas, los sistemas fragmentados, la escasez de profesionales calificados y la falta de apoyo de la alta dirección comprometen la gobernanza de la información pública y la preservación de la memoria institucional (Guevara y Vega-Almeida, 2024; Mutsagondo y Khumalo, 2024). Una mala gestión puede generar lagunas en la documentación de las actividades gubernamentales, lo que afecta a la rendición de cuentas pública y la continuidad administrativa (Matlala y Ncube, 2025).

Estudios apuntan la necesidad de políticas claras, integración entre sistemas e inversiones en formación técnica (Rahman, 2020, 2021; Dorneles, Córrea y Flores, 2023; Ngoepe y Mello, 2021). La fragmentación de los sistemas, la falta de estandarización y la baja participación institucional dificultan una gestión eficaz. Asogwa, Ezeani y Asogwa (2021) destacan que las bibliotecas universitarias enfrentan problemas similares. Otros estudios destacan la importancia de una infraestructura tecnológica adecuada, políticas de archivos públicos, orientación formal de los archivos estatales y mecanismos legales (Mosweu y Bwalya, 2024; Mater y ahin, 2023; Malatji y Marutha, 2025). Finalmente, autores como Sem y Jindal (2024) refuerzan que la gestión de registros debe ser tratada como parte de la gobernanza de datos, integrando principios de procedencia, trazabilidad y preservación en la formulación de políticas de información pública.

### 3.2. PRESERVACIÓN DIGITAL

La literatura analizada revela que la preservación digital aún enfrenta desafíos importantes en el sector público, tanto desde el punto de vista técnico como institucional. Ntengenyane y Masenya (2022) destacan que la ausencia de políticas de información claras, sumada a la falta de compromiso institucional y la baja capacitación técnica, compromete la efectividad de las acciones de preservación. De manera similar, Jacobs y Lemekoana (2021) señalan las brechas en la capacitación del personal y la resistencia al cambio como factores que debilitan los procesos de preservación en los archivos judiciales digitales.

El estudio de Chikomba, Rodrigues y Ngoepe (2021) revela que la mayoría de las empresas paraestatales carecen de políticas formales para la preservación de registros digitales, lo que conduce a un manejo inadecuado de la información y a riesgos para la integridad de los documentos. Para los autores, la falta de infraestructura y el bajo conocimiento archivístico de los empleados agravan este escenario. Asogwa, Ezeani y Asogwa (2021) identifican que, en las bibliotecas universitarias, la falta de sistemas de gestión electrónica estandarizados compromete la preservación y la eficiencia organizacional.

Mosweu y Bwalya (2024) afirman que los sistemas utilizados no cumplen con los requerimientos mínimos de archivo, con fallas de seguridad, falta de copias de seguridad y pérdida de datos. Los autores abogan por la creación de archivos digitales nacionales como estrategia de política pública. Mater y ahin (2023) añaden que la transferencia de registros digitales a instituciones de archivo se ve obstaculizada por la falta de infraestructura técnica y orientación legal. Netshakhuma (2020, 2025) señala la falta de evaluación y estrategias claras como obstáculos para garantizar la autenticidad y la preservación a largo plazo de los registros digitales.

### 3.3. SEGURIDAD DE LOS REGISTROS DIGITALES

La seguridad de los registros digitales es uno de los temas centrales en la literatura reciente sobre gestión de archivos. Los estudios analizados abordan la integridad, la autenticidad y

la protección de los registros en entornos informáticos, destacando amenazas como la obsolescencia tecnológica, el error humano y los ciberataques.

Anand y Singh (2022) propusieron un método de autenticación para registros digitales en el área de la salud basado en *hashing* criptográfico y compresión de datos, capaz de garantizar la integridad de la información sin comprometer el rendimiento del sistema. Según los autores, el mecanismo permite detectar inmediatamente cualquier modificación en los datos, garantizando la trazabilidad y la fiabilidad de los registros digitales.

Shekgola y Ngoepe (2025) clasifican los riesgos a la seguridad de los registros digitales en tres grandes grupos: obsolescencia tecnológica, error humano y ciberataques. Para mitigar estos riesgos, sugieren el uso de migración y emulación de formatos, mecanismos de autenticación robustos y medidas como límites de tiempo de sesión en los sistemas de acceso. Afirman que «la detección y prevención de actividades no autorizadas depende de actualizaciones constantes y la mejora continua de los controles de seguridad».

Flynn y Olukoya (2025) proponen un enfoque técnico basado en la detección automatizada de actividades maliciosas en registros digitales, utilizando algoritmos de coincidencia aproximada combinados con aprendizaje automático. Los autores demuestran cómo los algoritmos *hash* pueden representar con precisión los datos de registro, lo que permite la identificación de comportamiento sospechoso y refuerza el papel de los registros digitales en la ciberseguridad organizacional.

St-Hilaire (2023) analiza la influencia de las estrategias de protección de datos en la gobernanza del riesgo digital en empresas estatales. Su investigación resalta la importancia de los sistemas de monitoreo, mecanismos de control y capacitación de los equipos de prevención del fraude. El autor sostiene que «la protección de los registros digitales se considera una inversión esencial en las empresas estatales» y que las prácticas innovadoras de mantenimiento de registros pueden reducir significativamente la exposición al riesgo digital.

Finalmente, Xiao, Xu y Liu (2021) destacan las debilidades estructurales en las instituciones que todavía utilizan métodos tradicionales de almacenamiento digital. Señalan la falta de normas nacionales coherentes, los recursos técnicos y humanos limitados y la falta de una cultura de seguridad institucional como factores que comprometen la preservación a largo plazo. Los autores afirman que «la ausencia de estándares unificados y la escasez de profesionales cualificados dificultan la implementación efectiva de medidas de seguridad digital».

Estos estudios demuestran que, si bien existen avances tecnológicos y propuestas innovadoras, la seguridad de los registros digitales aún depende de la consolidación de políticas institucionales, inversiones continuas y capacitación profesional para convertirse en una práctica consolidada y efectiva.

### 3.4. INTELIGENCIA ARTIFICIAL APLICADA A LA GESTIÓN DE REGISTROS

La literatura reciente destaca el potencial de la inteligencia artificial (IA) para transformar profundamente el ciclo de vida de los registros digitales, con aplicaciones que van desde la creación y la clasificación hasta el almacenamiento, la recuperación y la eliminación de

documentos. Sin embargo, los estudios también destacan límites y riesgos que exigen enfoques críticos y políticas específicas.

Modiba (2025) sostiene que la IA y las máquinas robóticas son herramientas poderosas con la capacidad de mejorar significativamente todas las etapas de la gestión de registros. El autor propone un marco rector que incorpora herramientas de IA en las actividades de archivo, destacando que estas tecnologías pueden realizar eficazmente prácticas como digitalización, indexación y eliminación, contribuyendo a la preservación, la seguridad y el acceso autorizado a los registros digitales.

El estudio de Svärd *et al.* (2024) señala que todavía hay poca investigación sobre los impactos de la IA en la gestión de registros críticos en las administraciones públicas, y destaca que el entorno regulatorio sobre el tema varía ampliamente entre países. Para los autores, la utilidad de la IA depende de la competencia técnica, una infraestructura digital robusta y prácticas de preservación de datos alineadas con estándares éticos y legales.

Modiba (2025), en un trabajo complementario, aboga por la creación de políticas institucionales específicas para orientar el uso de IA en la gestión de registros, incluyendo pautas de seguridad, respaldo y uso de sistemas integrados. Según el autor, estas medidas son fundamentales para garantizar la adopción segura, eficiente y ética de estas tecnologías en entornos de archivo.

Cushing y Osti (2023) abordan la IA como un agente de transformación en la percepción de la experiencia archivística digital. Argumentan que la tecnología debe ser tratada como una herramienta de apoyo, y no un reemplazo, del trabajo de archivo humano, y exigen una reflexión crítica sobre el papel de las habilidades profesionales en la era digital. Según los autores, la IA puede generar nuevas formas de caracterizar, comunicar y valorar el conocimiento archivístico en contextos digitales.

Siguiendo en este campo, Jaillant y Rees (2023) refuerzan la necesidad de establecer pautas éticas claras y fortalecer la comunicación interinstitucional como prerequisites para la implementación responsable de la IA en entornos de archivo. Nockels, Gooding y Terras (2024) abordan específicamente el uso de IA en tecnologías de reconocimiento de texto manuscrito (HTR), demostrando su impacto positivo en el acceso y la organización de colecciones documentales históricas, siempre que esté guiado por principios colaborativos y éticos.

Estos estudios convergen en la defensa de un modelo de adopción de la inteligencia artificial que integre la innovación tecnológica con los fundamentos de la archivística, evitando la descontextualización de los registros y preservando los valores de autenticidad, fiabilidad y responsabilidad documental.

### 3.5. LOS REGISTROS DIGITALES COMO PRUEBA

Los registros digitales desempeñan un papel central en la construcción de evidencias administrativas, jurídicas e históricas, y la literatura reciente analiza los desafíos y las estrategias para asegurar su autenticidad, integridad y cadena de custodia. Los estudios analizados defienden que, ante la creciente complejidad de los registros digitales y la volatilidad de los

entornos tecnológicos, es necesario revisar los modelos archivísticos tradicionales y adoptar enfoques capaces de lidiar con artefactos informacionales dinámicos y distribuidos.

McLeod y Lomas (2023) proponen una reconceptualización del registro digital a partir de la metáfora del «registro de DNA», caracterizándolo como un artefacto complejo con potencial de replicación y múltiples capas de metadatos. Para las autoras, es esencial superar la noción de documento fijo y adoptar modelos que reconozcan a los registros como entidades interactivas y evolutivas. En este sentido, sugieren «un modelo conceptual de registro digital que sea suficientemente robusto para capturar las capas de datos, metadatos y contexto» (McLeod y Lomas, 2023, p. 4).

Duranti, Rogers y Thibodeau (2022) refuerzan la centralidad de la procedencia y de la diplomática en la preservación de la autenticidad en entornos digitales. Según los autores, «la autenticidad de un documento archivístico digital está enraizada en su procedencia e integridad, y debe ser preservada mediante prácticas continuas de control y documentación de las transformaciones» (Duranti, Rogers y Thibodeau, 2022, p. 90), destacando que la autenticidad no es estática, sino construida a lo largo del ciclo de vida documental.

La perspectiva forense también gana protagonismo. Ries (2022) argumenta que los métodos forenses digitales amplían la capacidad de análisis crítico de los registros digitales, permitiendo la lectura de metadatos ocultos, sellos de tiempo, historiales de modificación y datos residuales. Para el autor, «el enfoque forense permite examinar no solo los contenidos visibles de un archivo digital, sino también los vestigios técnicos que componen su materialidad informacional» (Ries, 2022, p. 13). Esta perspectiva es esencial para garantizar el valor probatorio de los documentos nativos digitales.

Pan y Duranti (2020) discuten la admisibilidad de evidencias digitales en el sistema judicial, destacando que la ausencia de estructuras archivísticas compromete la confiabilidad de los registros presentados como prueba. Las autoras afirman que «en muchos casos, los registros digitales presentados como prueba no son gestionados por sistemas archivísticos adecuados, lo que compromete su credibilidad» (Pan y Duranti, 2020, p. 149).

Aun así, sostienen que es posible asegurar su confiabilidad siempre que los procesos de creación y mantenimiento sean documentados y auditables.

Lawan y Henttonen (2024, 2025) examinan el papel de los registros digitales en investigaciones anticorrupción, resaltando que la autenticidad y la accesibilidad de estos documentos son esenciales para garantizar la rendición de cuentas institucional. Según los autores, «los registros electrónicos autenticados y gestionados adecuadamente son esenciales para mantener la integridad de las evidencias en investigaciones anticorrupción» (Lawan y Henttonen, 2025, p. 2).

En una perspectiva más tecnológica, Gheorghiu (2023) explora la aplicación de los NFT como mecanismos de evidencia documental, destacando su uso como registro de autenticidad basado en blockchain. La autora define los NFT como «un registro basado en blockchain de la autenticidad de los derechos sobre un objeto único que contiene metadatos sobre el objeto que representa» (Gheorghiu, 2023, p. 91), y discute las implicaciones jurídicas y fiscales de esta estructura en procesos de validación de registros digitales.

Como se ha demostrado, los registros digitales como evidencias exigen enfoques multidisciplinarios y metodologías que combinen fundamentos archivísticos, herramientas tecnológicas y prácticas forenses. El mantenimiento de la cadena de custodia y de la integridad informacional es decisivo para su admisibilidad y confiabilidad en procesos jurídicos, administrativos e históricos.

### 3.6. CAPACITACIÓN EN GESTIÓN DE REGISTROS

La formación profesional para trabajar en la gestión de registros digitales es un tema recurrente en los estudios analizados, que resaltan las brechas entre los currículos académicos y las demandas contemporáneas del campo archivístico. Los estudios indican que, si bien muchos profesionales tienen formación formal en gestión documental, la mayoría no recibe formación específica sobre los aspectos técnicos y tecnológicos de la gestión de registros digitales.

Katuu (2024) observa que los currículos existentes «carecen de coherencia, contenidos prácticos y actualización tecnológica» (p. 337), lo que dificulta preparar adecuadamente a los profesionales para afrontar los retos que imponen los entornos digitales. Ngoepe, Jacobs y Moja-pelo (2024) complementan este diagnóstico identificando una «brecha crítica en las habilidades digitales entre los profesionales recién graduados», en parte debido a la omisión del tema de la gestión de registros digitales en los planes de estudio de la mayoría de las instituciones. Los autores recomiendan la inclusión de contenidos como diplomacia de archivos, investigación forense digital, gobernanza de la información, curación digital y seguridad de la información.

Mutsagondo y Khumalo (2023) también destacan la escasez de habilidades digitales, señalando una «preocupante escasez de habilidades electrónicas entre el personal responsable de la gestión de registros e información» (p. 2). La investigación reveló que los currículos académicos aún no se ajustan a las demandas del mercado laboral, siendo necesario actualizar los planes de estudio e invertir en formación continua. Los autores argumentan que se debe ampliar la alineación entre las instituciones educativas y las demandas del sector público para garantizar una mayor eficiencia en la gestión digital.

Daniel y Ndumbaro (2023) se centran en las capacidades humanas orientadas a la preservación digital y muestran que las instituciones «aún carecen de áreas de planes de recursos humanos, concienciación, comprensión, habilidades, conocimientos y especialización técnica» (p. 15). Los autores recomiendan el establecimiento de planes de capacitación, contratación de especialistas y capacitación continua como medidas fundamentales para garantizar prácticas sostenibles de conservación de registros digitales.

Finalmente, Huda (2023) amplía la discusión abordando los principios éticos que deben guiar la formación profesional en la gestión de registros digitales. El autor propone un enfoque que articula la competencia técnica y la sensibilidad ética, afirmando que «la gestión de registros digitales requiere no sólo conocimientos técnicos, sino también sensibilidad ética para abordar los impactos sociales y culturales de la información» (p. 7). Este equilibrio, según Huda, es esencial para formar profesionales capaces de responder a los retos de la era digital con responsabilidad y conciencia social.

### 3.7. SISTEMAS DE GESTIÓN DE REGISTROS EN EL SECTOR PÚBLICO

La literatura reciente indica que la adopción de sistemas de gestión de registros en el sector público enfrenta obstáculos técnicos, institucionales y regulatorios que comprometen su efectividad. La coexistencia de sistemas híbridos con documentos digitales y digitalizados, la falta de integración entre los sistemas transaccionales y las prácticas archivísticas y la fragilidad de las políticas institucionales comprometen la trazabilidad, la autenticidad y la confiabilidad de los registros digitales.

Manase, Mwantimwa y Ndimbwa (2025) analizaron el uso de estos sistemas en instituciones públicas e identificaron que gran parte de los registros son resultado de la digitalización de documentos originalmente en papel. El uso eficaz de los sistemas depende de una infraestructura de TIC adecuada, el apoyo de la alta dirección, la existencia de marcos jurídicos claros, personal cualificado y formación de los usuarios.

Mosweu y Ngoepe (2021) destacan que, si bien los sistemas de contabilidad gubernamental contribuyen a la eficiencia informativa, la falta de integración con las prácticas archivísticas compromete la producción de registros confiables. Según los autores, la confiabilidad depende no solo de las funcionalidades técnicas, sino también de la implementación sistemática de políticas de gestión documental. Advierten que «la ausencia de un enfoque archivístico en los sistemas de contabilidad gubernamental pone en riesgo la integridad de los registros digitales, lo que impacta directamente en la rendición de cuentas y la gobernanza».

Complementando esta discusión, Klareld (2022) introduce el concepto de informática de mantenimiento de registros, proponiendo una visión más sistémica de la interacción entre tecnologías, registros y prácticas organizacionales. Para el autor, este enfoque permite comprender la creciente complejidad de los registros digitales y la necesidad de adaptar los sistemas a las dinámicas institucionales e informacionales contemporáneas.

A su vez, Oktarina *et al.* (2023) analizan los sistemas de gestión de registros en el contexto educativo y su relación con la rendición de cuentas y la transparencia institucional. El estudio concluye que dichos sistemas son esenciales para garantizar la rendición de cuentas de las escuelas, ya que «gestionar los documentos escolares equivale a proteger los intereses de los directores, docentes, estudiantes, padres y otras partes interesadas» (p. 25).

Estos estudios demuestran que la efectividad de los sistemas de gestión de documentos en el sector público depende no solo de soluciones tecnológicas, sino también de políticas integradas, formación profesional continua y coordinación institucional capaces de asegurar prácticas archivísticas consistentes y alineadas con la transformación digital del Estado.

### 3.8. REGISTROS E INFORMACIÓN DIGITALES

Esta categoría reúne estudios que exploran los registros digitales como construcciones informativas y sociales. Las obras analizan cómo el acceso, la representación y el uso de los registros digitales reflejan dinámicas de poder, procesos de exclusión y disputas narrativas, yendo más allá de una concepción puramente técnica o burocrática de los documentos de archivo.

Jaillant (2022) analiza las principales barreras que dificultan el acceso a los archivos digitales digitalizados y nativos, clasificándolas en tres categorías: legales (como las restricciones de derechos de autor y protección de datos), técnicas (falta de interoperabilidad e infraestructura inadecuada) e institucionales/culturales (como la ausencia de políticas claras y la resistencia organizacional). Como soluciones, propone alianzas entre archivos, investigadores y comunidades, además del desarrollo de políticas institucionales que equilibren la ética, la privacidad y la apertura de la información.

La libertad de información es un tema central en el trabajo de Mojapelo (2024), quien define este derecho como «el acceso de las personas y organizaciones a la información pública en poder de las entidades gubernamentales». El autor sostiene que los registros digitales son objetos fundamentales para la implementación de la transparencia y la rendición de cuentas, especialmente en la lucha contra la corrupción y en la promoción de la gobernanza democrática basada en datos abiertos y accesibles.

Lapp (2023) propone un enfoque alternativo a la neutralidad archivística tradicional, introduciendo el concepto de fabulación provenzal, definido como una forma de escritura archivística que incorpora afectos, lagunas y posibilidades no realizadas. El autor sostiene que la fabulación de archivo permite el surgimiento de subjetividades silenciadas, desafiando las convenciones positivistas de veracidad documental. Además, utiliza el concepto de difracción archivística como método para analizar los cambios de significado en diferentes contextos, visibilizando relaciones, ausencias y tensiones en la producción de registros.

En una línea similar, Özdemir e İçimsoy (2021) investigan cómo las prácticas institucionales y las políticas de acceso contribuyen al silenciamiento de las comunidades marginadas en los archivos públicos turcos. Los autores denuncian la invisibilidad de ciertas voces sociales y defienden la reinención de los archivos como espacios plurales e inclusivos, capaces de romper con el monopolio estatal de la narrativa histórica. Proponen que la gestión de la información incorpore principios de justicia social y representación equitativa de la diversidad histórica y cultural.

Estos estudios demuestran que los registros digitales son también artefactos simbólicos y políticos, cuya organización y acceso reflejan disputas sobre la memoria, la verdad y la ciudadanía. La gestión de archivos, por tanto, debe considerar los aspectos técnicos y sociales de la información, garantizando no solo la preservación, sino también la pluralidad y la accesibilidad de los registros en el entorno digital.

### 3.9. REPRESENTACIÓN DE REGISTROS DIGITALES

La representación de registros de archivos digitales es uno de los campos más complejos y desafiantes de la archivística contemporánea. La literatura analizada aborda tanto los fundamentos teóricos como los aspectos prácticos involucrados en la descripción, la organización y la vinculación de estos registros en entornos digitales dinámicos y distribuidos.

Trace (2020) problematiza la aplicación de modelos jerárquicos tradicionales en contextos digitales, afirmando que «la estructura jerárquica de ordenamiento y descripción a menudo

entra en conflicto con los modos digitales de producción y uso de la información» (p. 345). El autor propone enfoques alternativos, como modelos de procedencia múltiple y de red, argumentando que «la lógica del entorno digital desafía los principios archivísticos clásicos, exigiendo estructuras descriptivas que reflejen conexiones no lineales» (p. 349). Además, sostiene que «romper con la neutralidad implica considerar los contextos sociales y políticos al construir descripciones» (p. 351).

Bak (2024) profundiza el debate sobre la procedencia digital, defendiéndola como la «historia completa del registro», que incluye a múltiples actores, desde diseñadores y creadores de tecnología hasta archivistas. El autor destaca la importancia de la materialidad digital y señala que la procedencia debe considerar «cuándo el contenido es generado por IA y cuándo no», citando incluso requisitos recientes de la legislación estadounidense (Orden Ejecutiva 14110/2023).

La cuestión de la materialidad también es abordada por Oliveira, Kaya y Roncaglio (2022), quienes discuten la constitución de la información como fenómeno social, técnico y cultural. Para los autores, «la producción, circulación y consumo de información están permeados por diversos elementos materiales» (p. 9), cuestionando la idea de que los registros digitales serían inmateriales y destacando la infraestructura técnica y los procesos algorítmicos involucrados.

Henntonen y Packalén (2024) introducen el concepto de «dispersión digital» para describir un fenómeno en el que la información existe simultáneamente en múltiples formatos y plataformas, lo que dificulta su descripción y gestión. Según los autores, una representación efectiva debe abordar este desafío articulando estrategias que aborden la fragmentación y la redundancia de los datos.

Allam (2023) aporta una perspectiva más técnica, defendiendo el uso del estándar EAD (Encoded Archival Description) como el más adecuado para representar registros digitales, debido a su compatibilidad con sistemas automatizados y amplia adopción internacional. Katuu (2024) destaca la importancia de las representaciones ontológicas y los modelos de madurez, como DPCMM y ECM3, para evaluar la eficacia y la preparación institucional en la gestión de la información de archivo digital.

El trabajo de Campos, Rondinelli y Campos (2022) propone una concepción ontológica del soporte archivístico digital, tratado como una «entidad relacional» que soporta el contenido, la estructura y el contexto. Para los autores, la falta de definiciones precisas sobre el soporte compromete la interoperabilidad semántica y la representación precisa en los sistemas de información.

Finalmente, Lueg y Twidale (2020) cuestionan la representación del «individuo digitalizado», señalando que los datos no capturan de forma neutral la realidad de las personas. Para los autores, los registros digitales son fragmentos construidos mediante elecciones de codificación, y afirman que «ninguna cantidad de datos puede capturar completamente el significado que una acción tiene para un individuo» (p. 605), advirtiendo de los límites éticos e interpretativos de la representación digital de la experiencia humana.

#### 4. CONCLUSIONES

Esta revisión sistemática de la literatura, realizada con trabajos científicos publicados entre 2020 y 2025, destaca un cambio significativo en el enfoque de la investigación sobre registros de archivos digitales en comparación con la revisión anterior, publicada en 2019<sup>1</sup>. En ese momento, el principal foco de preocupación estaba en los aspectos estructurales de la descripción de archivos digitales, con énfasis en la estandarización de metadatos, modelos de representación conceptual y la interoperabilidad de sistemas como bases para garantizar la autenticidad y la preservación de los registros digitales.

En la presente revisión se observa una importante ampliación del alcance temático, con énfasis en las dimensiones prácticas, políticas e institucionales de la gestión y el uso de registros digitales. Las preocupaciones sobre la estructura descriptiva siguen siendo relevantes, como se señaló en la categoría sobre representación, pero ahora aparecen integradas en un contexto más complejo, en el que la archivística está directamente conectada con los desafíos de la transformación digital, la gobernanza de la información, la transparencia pública y las tecnologías emergentes.

Temas como la gestión de registros digitales en el sector público, la seguridad de la información, el uso de inteligencia artificial, la valorización de los registros como evidencia documental y los debates sobre la formación profesional y los sistemas de gestión digital son centrales en los estudios analizados. Estos temas indican una creciente preocupación por la eficacia de los procesos de archivo frente a la complejidad informativa y tecnológica del presente.

Además, la incorporación de enfoques críticos en los estudios sobre registros e información digitales y su representación demuestra una ciencia archivística que se esfuerza no solo por mantenerse al día con las innovaciones técnicas, sino también por reflexionar sobre los impactos sociales, culturales y éticos de la producción, el uso y el control de registros en entornos digitales.

En resumen, esta revisión revela que la investigación archivística internacional ha avanzado desde el debate técnico-descriptivo a un campo más integral y estratégico, centrado en la gestión integrada de los documentos digitales, su uso como activo institucional y su función pública como instrumento de memoria, evidencia y derecho. La consolidación de este nuevo campo de interés requiere, a su vez, el fortalecimiento de las políticas públicas, la articulación entre teoría y práctica y la ampliación de la formación crítica y técnica de los profesionales archivistas frente a los desafíos de la era digital.

1 El artículo completo con los resultados de la revisión anterior está disponible para su descarga gratuita en el sitio de la Revista Brasileña de Preservación Digital: <https://econtents.sbu.unicamp.br/inpec/index.php/rebpred/article/view/15890>

## REFERENCIAS BIBLIOGRÁFICAS

- Asogwa, B. E., Ezeani, C. N. y Asogwa, M. N. (2021). Status of electronic records management (e-RM) in African university libraries: Experience from Nigerian universities. *Library Management*, 42(8/9), 515–530. <https://doi.org/10.1108/LM-04-2021-0036>
- Bak, G. (2024). Digital provenance. *Archival Science*, 24(4), 847–869. <https://doi.org/10.1007/s10502-024-09462-w>
- Campos, L. M., Rondinelli, R. C. y Campos, M. L. D. (2022). O suporte do documento arquivístico digital: Uma proposta de definição conceitual apoiada nos princípios da teoria do conceito e da ontologia formal. *Acervo*, 35(2), 23. <https://cip.brapci.inf.br/download/203728>
- Chikomba, A., Rodrigues, A. y Ngoepe, M. (2021). Management of digital records in selected financial services parastatals in Zimbabwe. *Information Development*, 37(2), 245–257. <https://doi.org/10.1177/0266666920906858>
- Cushing, A. L. y Osti, G. (2023). «So how do we balance all of these needs?»: How the concept of AI technology impacts digital archival expertise. *Journal of Documentation*, 79(7), 12–29. <https://doi.org/10.1108/JD-08-2022-0170>
- Daniel, J. y Ndumbaro, F. (2023). An assessment of human resource capabilities in supporting digital records preservation: A case of RAMD and RITA, Tanzania. *Records Management Journal*, 33(1), 1–19. <https://doi.org/10.1108/RMJ-02-2021-0003>
- Dorneles, S. L., Côrrea, R. F. y Flores, D. (2023). Subsídios para política pública de gestão arquivística de documentos digitais: Revisão sistemática de literatura. *Perspectivas em Ciência da Informação*, 28. <https://doi.org/10.1590/1981-5344/46277>
- Duranti, L., Rogers, C. y Thibodeau, K. (2022). Authenticity. *Archives & Records*, 43(2), 188–203. <https://doi.org/10.1080/23257962.2022.2054406>
- Flynn, R. y Olukoya, O. (2025). Using approximate matching and machine learning to uncover malicious activity in logs. *Computers and Security*, 151, 104312. <https://doi.org/10.1016/j.cose.2025.104312>
- Gheorghiu, L. (2023). Policy forum: Non-fungible tokens and their income tax treatment. *Canadian Tax Journal*, 71(1), 83–99. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4435860](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4435860)
- Guevara, J. D. y Vega-Almeida, R. (2024). The nexus between records management and digital government in Cuba: Analysis of the regulatory framework. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 29. <https://doi.org/10.5007/1518-2924.2024.e96244>
- Henttonen, P. y Packalén, S. (2024). Results of archival appraisal: A study of a Finnish City. *Archival Science*, 24(1), 21–39. <https://doi.org/10.1007/s10502-023-09426-6>
- Huda, M. (2022). Empowering professional and ethical balance in digital record management. *Organizational Cybersecurity Journal: Practice, Process and People*, 2(1), 60–73. <https://doi.org/10.1108/O CJ-06-2021-0016>
- Jacobs, L. y Lemekoana, T. (2021). Electronic records management at the Department of Justice and Constitutional Development, South Africa: Towards e-government. *Mousaion*, 39(4), 1–16. <http://doi.org/10.25159/2663-659X/9967>
- Jaillant, L. (2022). How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Archival Science*, 22(3), 417–436. <https://doi.org/10.1007/s10502-022-09390-7>
- Jaillant, L. y Rees, A. (2023). Applying AI to digital archives: Trust, collaboration and shared professional ethics. *Digital Scholarship in the Humanities*, 38(2), 571–585. <https://doi.org/10.1093/llc/fqac073>

- Katuu, S. (2024). Advancing archives and records management professionals in Africa. *Global Knowledge, Memory and Communication*, 73(3), 331–353. <https://doi.org/10.1108/GKMC-05-2022-0100>
- Klareld, A.-S. (2021). A reliable reflection? Challenges when documenting physical infrastructure. *Records Management Journal*, 31(2), 134–157. <https://doi.org/10.1108/RMJ-11-2020-0040>
- Lapp, J. M. (2023). «The only way we knew how»: Provenancial fabulation in archives of feminist materials. *Archival Science*, 23(1), 117–136. <https://doi.org/10.1007/s10502-021-09376-x>
- Lawan, A. A. y Henttonen, P. (2024). Preserving evidence integrity: the key to efficient anti-corruption investigations. *Records Management Journal*, 34(1), 41–57. <https://doi.org/10.1108/RMJ-03-2023-0016>
- Lawán, A. A. y Henttonen, P. (2025). Shaping anti-corruption strategies: Investigator perspectives on electronic records. *Journal of Financial Crime*, 32(3), 558–571. <https://www.emerald.com/insight/content/doi/10.1108/jfc-03-2024-0120/full/html>
- Lueg, C. P. y Twidale, M. B. (2020). What data don't tell you about the 'digitized individual'. *Library Trends*, 68(4), 599–610. <https://doi.org/10.1353/lib.2020.0013>
- Malatji, O. y Marutha, N. S. (2025). Implementation of legislative framework governing records management throughout the life cycle in the Limpopo provincial government of South Africa. *Global Knowledge, Memory and Communication*, 74(1/2), 253–271. <https://www.emerald.com/insight/content/doi/10.1108/gkmc-08-2022-0195/full/html>
- Manase, J., Mwantimwa, K. y Ndimbwa, T. (2025). Use of records management systems in Tanzania public sector organisations. *Records Management Journal*, 35(1), 35–58. <https://www.emerald-com.ez15.periodicos.capes.gov.br/insight/content/doi/10.1108/rmj-12-2021-0041/full/html>
- Mater, E. C. y ahin, L. Ö. (2023). An examination of the challenges and awareness presented by the transfer of born-digital records in the public sector: A case study of Türkiye. *Information Development*. <https://doi.org/10.1177/02666669231214529>
- Matlala, M. E. y Ncube, T. R. (2025). Electronic records management amidst the seismic shift in the dynamic infosphere. *Records Management Journal*, 35(1), 59–74. <https://www.emerald.com/insight/content/doi/10.1108/rmj-04-2023-0022/full/html>
- McLeod, J. y Lomas, E. (2023). Record DNA: Reconceptualising digital records as the future evidence base. *Archival Science*, 23(3), 411–446. <https://doi.org/10.1007/s10502-023-09414-w>
- Modiba, M. (2025). Adoption of artificial intelligence to enhance records management practices at Gauteng Department of Education in South Africa. *Collection and Curation*, 44(1), 9–17. <https://www.emerald.com/insight/content/doi/10.1108/cc-12-2023-0044/full/html>
- Mosweu, O. y Bwalya, K. J. (2024). The challenges of post custodial management of digital records in Botswana laid bare. *Information Development*, 40(2), 266–281. <https://doi.org/10.1177/02666669221114867>
- Mosweu, O. y Ngoepe, M. (2021). Trustworthiness of digital records in government accounting system to support the audit process in Botswana. *Records Management Journal*, 31(1), 89–108. <https://doi.org/10.1108/RMJ-11-2019-0069>
- Mutsagondo, S. y Khumalo, N. B. (2024). An assessment of senior management support for the records and information management function in Zimbabwe's central government. *Collection and Curation*, 43(3), 86–93. <http://dx.doi.org/10.1108/CC-06-2023-0021>
- Ngoepe, M., Jacobs, L. y Mojapelo, M. (2024). Inclusion of digital records in the archives and records management curricula in a comprehensive open distance e-learning environment. *Information Development*, 40(2), 190–201. <https://doi.org/10.1177/02666669221081812>

- Ngoepe, M. y Mello, V. (2021). Integration of records management systems at a South African water utility company. *Global Knowledge, Memory and Communication*, 70(8–9), 801–816. <https://doi.org/10.1108/GKMC-08-2020-0125>
- Netsakhuma, N. S. (2020). Appraisal evaluation of records: The case of the University of Witwatersrand and the University of Venda. *Aslib Journal of Information Management*, 72(4), 489–508. <https://doi.org/10.1108/AJIM-02-2020-0043>
- Netsakhuma, N. S. (2025). The assessment of the appraisal of records: Case of the University of the Witwatersrand and the University of Venda. *Aslib Journal of Information Management*, 77(1), 191–211. <https://www.emerald.com/insight/content/doi/10.1108/ajim-10-2019-0277/full/html>
- Nockels, J., Gooding, P. y Terras, M. (2024). The implications of handwritten text recognition for accessing the past at scale. *Journal of Documentation*, 80(7), 148–167. <https://doi.org/10.1108/JD-09-2023-0183>
- Oktarina, N., Purasani, H. N., Suryanto, E., Murwatiningsih, M. y Sehabuddin, A. (2023). Constructing a conceptual electronic record management system model based on eight Indonesian education standards to support school accountability. *Journal of Information Science Theory and Practice*, 11(3), 16–28. <https://www.koreascience.kr/article/JAKO202327052459296.pdf>
- Oliveira, K. V. R., Kaya, G. T. y Roncaglio, C. (2022). Ciências da informação e humanidades digitais: Produção, consumo e materialidade da informação em plataformas digitais. *Acervo*, 35(1). <https://brapci.inf.br/index.php/res/v/193731>
- Özdemir, L. y İçimsoy, O. (2021). Perceived silence in Turkish archives. En A. Gilliland, S. McKemmish y A. Lau (Eds.), *Pluralizing the archival curriculum in the multicultural world*. Routledge. <https://doi.org/10.4324/9781003003618-9>
- Pan, W. y Duranti, L. (2020). Sitting in limbo or being the flaming Phoenix: The relevance of the archival discipline to the admissibility of digital evidence in China. *Archives and Manuscripts*, 48(3), 300–327. <https://doi.org/10.1080/01576895.2020.1819162>
- Rahman, H. (2020). A review of digital records management needs for academic libraries. *Library Hi Tech News*, 37(3), 21–22. <https://doi.org/10.1108/LHTN-11-2019-0083>
- Rahman, M. H. (2021). The new trend for digital records management in university libraries of Bangladesh. *Library Philosophy and Practice*, (4740). <https://digitalcommons.unl.edu/libphilprac/4740>
- Ries, T. (2022). Digital history and born-digital archives: The importance of forensic methods. *Journal of the British Academy*, 10, 157–185. <https://doi.org/10.5871/jba/010.157>
- Sen, S. y Jindal, T. (2024). What does a record have to do with it? Re-situating records management within Indian public data governance and policy. *Data and Policy*, 6. <https://doi.org/10.1017/dap.2024.30>
- Shekgola, M. M. y Ngoepe, M. (2025). Ingesting digital archives into long-term storage system through free open-source software in South Africa. *Collection and Curation*, 44(1), 25–33. <https://www.emerald.com/insight/content/doi/10.1108/cc-02-2024-0003/full/html>
- St-Hilaire, W. A. (2023). Leading with digital technologies governance in the state-owned enterprises. *International Journal of Public Administration*, 46(2), 107–120. <https://doi.org/10.1080/01900692.2021.1993898>
- Svärd, P., Guerrero, E., Balogun, T., Saurombe, N., Jacobs, L. y Henttonen, P. (2024). Local regulations for the use of artificial intelligence in the management of public records – A literature review. *Records Management Journal*, 34(2/3), 109–130. <https://doi.org/10.1108/RMJ-10-2023-0061>
- Trace, C. B. (2020). Keeping records in context? Breaking the theory and practice of archival classification and arrangement. *The American Archivist*, 83(2), 322–372. <https://doi.org/10.17723/0360-9081-83.2.322>

## PRESERVACIÓN DIGITAL Y SU VINCULACIÓN CON LOS DATOS ABIERTOS DE REPOSITARIOS INSTITUCIONALES EN ESPAÑA: UN ANÁLISIS DESCRIPTIVO

### *DIGITAL PRESERVATION AND ITS LINKAGE TO OPEN DATA IN INSTITUTIONAL REPOSITORIES IN SPAIN: A DESCRIPTIVE ANALYSIS*

Kevin LEÓN-GAVILANEZ

*Departamento de Biblioteconomía y Documentación. Universidad de Salamanca. España*  
kfleong@usal.es

 <https://orcid.org/0000-0001-7936-7961>

**RESUMEN:** Este estudio analiza el estado actual de la preservación digital en los repositorios institucionales universitarios en España, con especial atención a su vinculación con los datos abiertos. Mediante una metodología descriptiva y de observación documental, se examinaron 50 repositorios activos pertenecientes a universidades públicas, considerando 13 dimensiones clave, como infraestructura tecnológica, tipología de contenidos, políticas institucionales, estándares aplicados, recursos disponibles y sostenibilidad proyectada. Los resultados revelan que, aunque el 70 % de los repositorios emplean plataformas técnicamente adecuadas como DSpace, solo el 24 % cuentan con políticas formales de preservación digital y menos del 35 % aplican estándares internacionales como OAIS o PREMIS. Asimismo, apenas un 30 % disponen de personal técnico especializado y solo el 40 % cuentan con financiación específica. Se identificaron buenas prácticas en repositorios como GREDOS (Universidad de Salamanca), e-Spacio (UNED) y Dialnet (Universidad de La Rioja), pero también una marcada desigualdad institucional. La discusión destaca que la preservación digital sigue siendo una función secundaria en muchas universidades, limitada por la falta de planificación estratégica, coordinación nacional y recursos estables. La proyección a 10 años indica que el 44 % de los repositorios presenta un nivel bajo de sostenibilidad. Se concluye que es urgente articular una política nacional de preservación digital universitaria, fortalecer las redes colaborativas, profesionalizar al personal y garantizar la sostenibilidad del conocimiento científico almacenado. La preservación digital debe integrarse plenamente en las políticas de ciencia abierta como componente estratégico e ineludible para garantizar la sostenibilidad, la trazabilidad, la accesibilidad del conocimiento a largo plazo ya no solo en entornos universitarios, sino también en el ámbito de la investigación.

PALABRAS CLAVE: preservación digital; repositorios institucionales; ciencia abierta; datos abiertos; universidades públicas; España.

ABSTRACT: This study analyzes the current state of digital preservation in institutional university repositories in Spain, with particular emphasis on its connection to open data. Using a descriptive methodology and documentary observation, 50 active repositories belonging to public universities were examined, considering 13 key dimensions such as technological infrastructure, content typology, institutional policies, applied standards, available resources, and projected sustainability. The results reveal that although 70 % of the repositories use technically suitable platforms such as DSpace, only 24 % have formal digital preservation policies, and fewer than 35 % apply international standards such as OAIS or PREMIS. Additionally, only 30 % have specialized technical staff, and just 40 % receive dedicated funding. Good practices were identified in repositories such as GREDOS (University of Salamanca), e-Spacio (UNED), and Dialnet (University of La Rioja), although significant institutional inequality was also observed. The discussion highlights that digital preservation remains a secondary function in many universities, hindered by the lack of strategic planning, national coordination, and stable resources. A ten-year projection shows that 44 % of repositories present a low level of sustainability. It is concluded that there is an urgent need to articulate a national policy for university digital preservation, strengthen collaborative networks, professionalize staff, and ensure the long-term sustainability of stored scientific knowledge. Digital preservation must be fully integrated into open science policies as a strategic and essential component to guarantee the sustainability, traceability, and accessibility of knowledge over time –not only within university environments but also across the broader research landscape, ensuring long-term value, institutional accountability, and equitable access to scientific information.

KEYWORDS: digital preservation; institutional repositories; open science; open data; public universities; Spain.

## 1. INTRODUCCIÓN

La preservación digital se ha consolidado como un componente esencial dentro de las políticas de gestión de la información científica y académica en el entorno universitario contemporáneo. En el contexto actual de transformación digital y de crecimiento sostenido de la ciencia abierta (Hallo Carrasco y De la Fuente, 2013), las universidades se ven enfrentadas al reto no solo de facilitar el acceso a la información mediante plataformas digitales, sino también de garantizar la integridad, la autenticidad, la accesibilidad y la disponibilidad a largo plazo de los recursos que albergan. En este marco, los repositorios institucionales no solo cumplen la función de almacenar la producción académica y científica, sino que también constituyen una pieza clave en la conservación del patrimonio digital de las instituciones de educación superior.

En los últimos años, el auge del acceso abierto ha transformado profundamente los modelos de publicación y diseminación del conocimiento. A través de políticas institucionales, mandatos gubernamentales e iniciativas internacionales como el Plan S, impulsado por la cOAlition S en Europa (Abadal *et al.*, 2019), se ha promovido el depósito obligatorio de los

resultados de investigación financiados con fondos públicos en repositorios de libre acceso (Barba-Salazar *et al.*, 2024). En España, esta transformación ha estado respaldada por normativas como la Ley 14/2011, de la Ciencia, la Tecnología y la Innovación (2011), que refuerza el compromiso del sistema de ciencia y tecnología con la apertura del conocimiento, incentivando a las universidades y centros de investigación a establecer mecanismos adecuados para el depósito, la conservación y la difusión de sus resultados científicos. Sin embargo, dicha normativa, si bien establece la necesidad del acceso abierto, no contempla en profundidad el conjunto de requerimientos necesarios para garantizar la preservación digital de los contenidos publicados (Azorín *et al.*, 2020; Ochoa-Gutiérrez *et al.*, 2021; Segundo Martín *et al.*, 2022; Castro Leal, 2024).

En este sentido, es importante subrayar que el concepto de acceso abierto no se limita únicamente a la disponibilidad inmediata de los documentos o *datasets*, sino que implica también la sostenibilidad de ese acceso a través del tiempo. La accesibilidad sin barreras debe estar acompañada de estrategias de conservación a largo plazo que aseguren que los archivos digitales no se pierdan, corrompan o vuelvan inaccesibles debido a fallos tecnológicos, obsolescencia de formatos o discontinuidades institucionales. La falta de estrategias específicas de preservación digital en los repositorios puede generar, en el mediano y largo plazo, una situación de vulnerabilidad en cuanto a la pérdida del patrimonio académico, especialmente en instituciones que no cuentan con infraestructuras consolidadas ni políticas claras de conservación (Library of Congress, 2008; Raventós Pajares, 2017; Suárez, 2024).

La distinción entre acceso y preservación resulta crucial. Un documento o conjunto de datos puede estar disponible hoy en repositorio digital, pero, si no se adoptan mecanismos de preservación técnica y organizativa, ese recurso corre el riesgo de volverse ilegible o inaccesible en el futuro. Los desafíos técnicos incluyen la migración de formatos obsoletos, la actualización de sistemas de gestión documental, la verificación de integridad de archivos y la documentación de metadatos que describan no solo el contenido, sino también su contexto técnico, legal y de uso. Estos desafíos no son triviales y exigen un enfoque integral que combine recursos tecnológicos, normativos y humanos (Ferrerías Torres, 2010).

A nivel internacional, han surgido diversas iniciativas orientadas a estandarizar y estructurar la preservación digital. El modelo OAIS (*Open Archival Information System*), adoptado como norma ISO 14721:2012, establece una arquitectura conceptual para la preservación y la recuperación de información digital a largo plazo. Este modelo ha sido ampliamente referenciado en instituciones como bibliotecas nacionales, archivos digitales y repositorios científicos. Asimismo, estándares como METS (*Metadata Encoding and Transmission Standard*) y PREMIS (*Preservation Metadata: Implementation Strategies*) permiten organizar y registrar información clave para la conservación digital, incluyendo detalles sobre el origen, los derechos, los eventos técnicos y el contexto de los objetos digitales. No obstante, su adopción en el entorno universitario español sigue siendo limitada y, en muchos casos, parcial.

La implementación de estos marcos conceptuales y tecnológicos depende, en gran medida, de la voluntad institucional, la disponibilidad presupuestaria, la capacitación del personal y la existencia de políticas claramente definidas. En España, aunque más del 90 % de las universidades públicas disponen de repositorios institucionales, la proporción que ha incorporado

estándares avanzados de preservación digital es significativamente menor. De acuerdo con datos del grupo de repositorios de la Red de Bibliotecas Universitarias (REBIUN) y los registros internacionales de ROAR (*Registry of Open Access Repositories*) y OpenDOAR (*Directory of Open Access Repositories*), el software más utilizado en estos repositorios es DSpace, una plataforma que soporta la inclusión de estándares, de preservación, pero cuya configuración avanzada requiere de conocimientos técnicos específicos y una inversión sostenida en mantenimiento y desarrollo (De-Filippo, 2023).

Adicionalmente, la preservación digital debe entenderse como un compromiso ético y social de las universidades. Al asegurar la permanencia de los contenidos científicos, las instituciones no solo protegen su memoria intelectual, sino que también garantizan el acceso equitativo a los saberes producidos con fondos públicos. Este compromiso con la sostenibilidad del conocimiento forma parte del contrato social de la universidad con la sociedad, especialmente en contextos donde la transparencia, la rendición de cuentas y la apertura del conocimiento son valores centrales (Batthyány, 2023). La falta de preservación no solo implica riesgos de pérdidas de información, sino que también puede comprometer la reproducibilidad de la ciencia, uno de los principios fundamentales del movimiento de ciencia abierta. Cuando los datos de investigación, los artículos y otros productos académicos no se preservan adecuadamente, se limita la posibilidad de validar resultados, reutilizar datos para nuevos estudios y construir conocimiento acumulativo. En este contexto, la preservación digital no es solo un aspecto técnico y administrativo, sino también un factor crítico para la calidad y la confiabilidad del sistema de investigación (Monteagudo-Haro y Prieto Gutiérrez, 2024).

Desde el punto de vista institucional, integrar la preservación digital dentro de la estrategia general de gestión del conocimiento requiere superar una serie de barreras estructurales. Entre ellas destacan la fragmentación de responsabilidades entre bibliotecas, servicios informáticos y grupos de investigación; la falta de formación especializada del personal encargado de la gestión de los repositorios; y la ausencia de marcos normativos internos que definan con claridad los procedimientos y las responsabilidades asociadas a la conservación digital. Además, la ausencia de indicadores específicos de calidad o auditorías periódicas sobre las prácticas de preservación contribuye a que esta dimensión sea, en muchos casos, relegada frente a otros aspectos más visibles o urgentes (Abad García *et al.*, 2022).

Cabe señalar que la preservación digital también se enfrenta a desafíos en cuanto a la sostenibilidad económica. Los costes asociados al almacenamiento seguro, la infraestructura tecnológica, la migración de formatos, el soporte técnico y la formación del personal son significativos. Sin mecanismos de financiación estables, las estrategias de preservación corren el riesgo de convertirse en prácticas puntuales, dependientes de proyectos temporales o de la iniciativa individual de ciertos equipos. En este sentido, sería conveniente que las agencias financiadoras de investigación, así como los gobiernos autonómicos y el Estado, integren en sus políticas de ciencia abierta criterios y recursos específicos para garantizar la preservación digital de la producción científica financiada públicamente (De-Giusti, 2023).

Este estudio parte, por tanto, de la necesidad de comprender en qué medida los repositorios institucionales universitarios en España están en condiciones de garantizar no solo el acceso abierto, sino también la conservación a largo plazo de los contenidos que custodian. Para

ello, se plantea un análisis descriptivo de las políticas, las prácticas y las tecnologías asociadas a la preservación digital, con especial atención a la vinculación entre la sostenibilidad del acceso abierto y la implementación de estrategias de conservación. La hipótesis de trabajo sugiere que, aunque existe una infraestructura digital ampliamente extendida, persisten importantes carencias en cuanto a la planificación, la estandarización y el financiamiento de la preservación digital en los repositorios universitarios.

El análisis de casos y la identificación de buenas prácticas permitirán caracterizar el panorama actual, establecer comparaciones entre distintas instituciones y proponer recomendaciones orientadas a mejorar la sostenibilidad y la calidad de la preservación digital. Al hacerlo, se espera contribuir al fortalecimiento de las capacidades institucionales en este ámbito, así como al diseño de políticas públicas más integradas y eficaces para asegurar la permanencia del conocimiento académico y científico producido en el ámbito universitario español.

Así, la preservación digital debe entenderse como un eje estratégico y transversal para asegurar la continuidad, la integridad y la proyección del ecosistema de ciencia abierta. Los repositorios institucionales, lejos de limitarse a funciones de archivo, representan instrumentos dinámicos para conservar, difundir y democratizar el conocimiento generado en el ámbito académico. Por ello, este estudio tiene como finalidad ofrecer un análisis fundamentado que contribuya al fortalecimiento de estas infraestructuras desde enfoques técnicos, organizativos y normativos, alineados con los principios del acceso abierto y el compromiso social de las universidades.

## 2. METODOLOGÍA

Para abordar el estudio de la preservación y su vinculación con los datos abiertos en los repositorios institucionales universitarios en España, se adoptó un enfoque metodológico descriptivo cualitativo e interpretativo, basado en la observación no participativa y en el análisis documental sistemático. Este enfoque permitió caracterizar la realidad observable de los repositorios institucionales sin intervenir en su funcionamiento, lo que resulta adecuado cuando el objetivo es examinar infraestructuras, políticas y estrategias preexistentes a partir de datos accesibles públicamente.

La elección de una metodología descriptiva responde a la necesidad de obtener un retrato detallado del estado actual de la preservación digital en el sistema universitario español. Dado que el campo de estudio implica una dimensión tecnológica, organizativa e institucional, el método descriptivo permite registrar y sistematizar las características y las condiciones reales de los objetos de análisis –los repositorios– sin necesidad de modificar su entorno o recurrir a la experimentación. Asimismo, la observación indirecta, centrada en el estudio de fuentes documentales disponibles en línea, fue considerada la estrategia más apropiada por razones de acceso, coherencia con el diseño del estudio y viabilidad ética.

El procedimiento metodológico se estructuró en varias etapas, combinando la recolección de datos mediante análisis documental con técnicas de análisis comparativo y codificación

cualitativa. En la fase inicial, se llevó a cabo la identificación y la delimitación de la unidad de análisis, compuesta por los repositorios institucionales de universidades públicas españolas. Esta selección se realizó con base en su inclusión en la Red de Bibliotecas Universitarias (REBIUN) y en su registro activo en directorios internacionales como ROAR (*Registry of Open Access Repositories*) y el OpenDOAR (*Directory of Open Access Repositories*). Esta combinación permitió establecer un universo fiable, representativo y actualizado, focalizado en instituciones que adopten políticas explícitas de acceso abierto y que disponen de plataformas activas y documentadas.

Se excluyeron del análisis los repositorios de universidades privadas y centros de investigación no universitarios, así como aquellos sin registros activos en los directorios internacionales o que no contaban con información accesible en sus sitios web institucionales. Esta decisión metodológica responde a la necesidad de mantener la coherencia y la comparabilidad del análisis, focalizando el estudio en una categoría homogénea de instituciones públicas con estructuras similares de gestión académica, financiación pública y responsabilidad normativa.

Una vez identificada la muestra, se desarrolló un protocolo de análisis documental, orientado a recopilar información estructurada y comparable entre los distintos repositorios. Este protocolo incluyó variables con la infraestructura tecnológica (plataforma utilizada, versión del software, personalización), el tipo de contenidos almacenados (artículos científicos, tesis doctorales, datos de investigación, materiales docentes, etc.), las políticas institucionales de preservación (documentos normativos, planes estratégicos, compromisos declarados) y los mecanismos técnicos implementados para garantizar la conservación digital (respaldo, migración de formatos, control de integridad, uso de metadatos, etc.).

Asimismo, se incluyeron variables relativas al uso de estándares internacionales reconocidos, tales como el modelo OAIS (*Open Archival Information System*), el esquema de metadatos METS (*Metadata Encoding and Transmission Standard*) y el estándar PREMIS (*Preservation Metadata Implementation Strategies*), así como la adopción de principios FAIR (*Findable, Accessible, Interoperable, Reusable*) y políticas de cumplimientos con iniciativas como el Plan S. Cada repositorio fue evaluado según la presencia o ausencia de estos elementos, así como el grado de detalle o desarrollo de los mismo cuando se encontraban disponibles.

La información se recopiló a través de la exploración directa de los portales web de los repositorios, la lectura de sus políticas institucionales y la consulta de documentos técnicos, informes de autoevaluación, memorias anuales o cualquier otra fuente primaria de libre acceso que ofreciera evidencia verificable. Cuando fue necesario, también se consultaron bases de datos académicas para contextualizar la información con literatura científica especializada en preservación digital y ciencia abierta (como LISTA o LISA).

El proceso de recolección de datos se dividió en tres fases operativas. En la primera fase, se procedió a la caracterización básica de los repositorios: se registraron sus datos generales (nombre, universidad, año de creación), la plataforma tecnológica empleada (por ejemplo, DSpace, EPrints o Fedora Commons) y el tipo de materiales que albergaban. Esta fase permitió establecer un panorama general de la infraestructura digital existente en las universidades públicas españolas y detectar patrones tecnológicos recurrentes.

En la segunda fase, se abordó el análisis cualitativo y comparativo de las políticas y estrategias de preservación digital. En este punto, se identificaron documentos estratégicos o técnicos que dieran cuenta de los planes de preservación a largo plazo, incluyendo aspectos como la migración de formatos, la redundancia de copias, la planificación de respaldo, la documentación de cambios y versiones, la formación del personal técnico y la asignación de recursos específicos. También se examinó si los repositorios contaban con mecanismos de control de integridad (*checksums*, *hashes*), auditoría o verificación periódica de datos.

La tercera fase estuvo orientada a la identificación de buenas prácticas y desafíos comunes. A través de una lectura crítica de los casos, se extrajeron ejemplos positivos de implementación, como en el caso de GREDOS (Universidad de Salamanca), que documenta su uso de estándares METS y PREMIS, o e-Spacio (UNED), que ofrece políticas explícitas de respaldo y conservación digital. Del mismo modo, se identificaron limitaciones frecuentes, como la falta de documentación formal, la ausencia de políticas públicas sobre preservación o la inexistencia de personal especializado.

Para el procesamiento de los datos, se diseñó una matriz de análisis cualitativo que permitió comparar repositorios según categorías comunes. Esta matriz permitió codificar y clasificar la información, facilitando su interpretación posterior. Si bien el estudio no se basa en un enfoque cuantitativo estricto, se utilizaron frecuencias relativas y porcentajes para representar la presencia o ausencia de determinadas prácticas, con el fin de visualizar patrones y niveles de implementación entre los repositorios observados.

Posteriormente, se aplicó una técnica de triangulación de fuentes, que consistió en contrastar los resultados obtenidos con otras investigaciones previas, informes institucionales, artículos académicos y documentos técnicos elaborados por organizaciones internacionales como la Unesco, la OpenAIRE o la Digital Preservation Coalition. Esta triangulación permitió contextualizar los hallazgos en un marco más amplio y aportar solidez a las conclusiones.

El enfoque adoptado también contempló una dimensión ética: dado que toda la información utilizada era de dominio público y se accedió sin interferir en los sistemas internos de las instituciones, no fue necesario recurrir a mecanismos de consentimiento ni a la validación ética mediante comités especializados. Aun así, se respetaron los principios de transparencia, rigurosidad y trazabilidad en todas las etapas del análisis.

En suma, la metodología empleada permitió desarrollar un estudio riguroso, sistemático y fundado en fuentes verificables. El carácter descriptivo y observacional del enfoque facilitó la comprensión de la situación actual de la preservación digital en los repositorios institucionales universitarios en España, destacando las oportunidades, las debilidades y los desafíos que enfrentan estas infraestructuras en el contexto de la ciencia abierta. La combinación de técnicas documentales, análisis comparativo y estudios de caso proporcionó una visión holística que no solo permite evaluar el estado presente, sino también orientar líneas de mejora a futuro.

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. RESULTADOS

En consonancia con la metodología adoptada, el análisis de los 50 repositorios institucionales pertenecientes a universidades públicas españolas ha permitido construir una versión sistemática y detallada del estado actual de la preservación digital en dichas infraestructuras. Los resultados obtenidos se organizan en 13 dimensiones clave: 1) infraestructura tecnológica y plataformas utilizadas, 2) tipologías documentales alojadas, 3) políticas y estrategias de preservación digital, 4) implementación de estándares y prácticas técnicas, 5) prácticas técnicas específicas, 6) financiación específica y recursos humanos, 7) casos de buenas prácticas identificadas, 8) análisis comparativo por tipología universitaria, 9) evolución temporal de las estrategias de preservación, 10) comparación con iniciativas internacionales, 11) implicaciones de la falta de preservación, 12) papel de las redes interinstitucionales: REBIUN y más allá y 13) proyección de sostenibilidad a 10 años.

##### 3.1.1. *Infraestructura tecnológica y plataformas utilizadas*

El primer aspecto analizado fue la plataforma tecnológica sobre la cual se desarrollan los repositorios institucionales. A continuación, se recogen en la Tabla 1 las plataformas de los repositorios examinados:

Tabla 1. Distribución de plataformas tecnológicas utilizadas en repositorios institucionales en España

Plataforma tecnológica	n.º de repositorios	%
DSpace	35	70,00
EPrints	7	14,00
Fedora Commons	5	10,00
Otras (desarrollos propios)	3	6,00

Fuente: Elaboración propia.

DSpace se consolida como la plataforma predominante debido a su amplia documentación, comunidad de desarrollo activa y compatibilidad con estándares como OAIS, METS y PREMIS. Su adopción masiva permite asumir una cierta homogeneidad técnica en cuanto a las posibilidades de integración de políticas de preservación. Por el contrario, los desarrollos propios –aunque adaptados a necesidades institucionales– presentan más heterogeneidad en cuanto a capacidades de mantenimiento y actualización tecnológica, lo cual puede afectar negativamente su sostenibilidad a largo plazo.

### 3.1.2. *Tipologías documentales alojadas*

En segundo lugar, se analizó la diversidad de contenidos almacenados en los repositorios. Los resultados se resumen en la Tabla 2:

Tabla 2. Tipología de contenidos alojados en los repositorios institucionales en España

Tipo de contenido	Repositorios que alojan	%
Artículos científicos	40	80,00
Tesis doctorales y trabajos académicos	38	76,00
Conjuntos de datos de investigación	23	46,00
Recursos educativos abiertos (REA)	16	32,00
Documentos patrimoniales e históricos	11	22,00

Fuente: Elaboración propia.

Como se observa, los repositorios cumplen una función básica de difusión de la producción académica textual (artículos y tesis), mientras que el alojamiento de conjuntos de datos de investigación y materiales educativos aún es limitado. Esta situación evidencia una concepción del repositorio centrada en la publicación y menos enfocada en la conservación de datos complejos, lo cual representa una limitación para avanzar hacia un ecosistema de ciencia verdaderamente abierta e interoperable.

### 3.1.3. *Políticas y estrategias de preservación digital*

El análisis de políticas institucionales explícitas de preservación digital que se muestra en la Tabla 3 reveló que solo 30 de 50 repositorios (60 %) disponen de algún tipo de documento que aborde la preservación a largo plazo. No obstante, al examinar estos documentos más de cerca, se encontró que en muchos casos las políticas son declarativas, sin detallar procesos técnicos, recursos asignados, mecanismos de revisión o métricas de evaluación.

Tabla 3. Estado de desarrollo de políticas institucionales de preservación digital en repositorios institucionales en España

Políticas de preservación digital	n.º de repositorios	%
Política formal y detallada	12	24,00
Política básica o declarativa	18	36,00
Sin política publicada o identificable	20	40,00

Fuente: Elaboración propia.

Esta situación genera un mapa desigual en cuanto a la protección efectiva del patrimonio digital universitario. Las instituciones con políticas robustas (por ejemplo, UNED o la Universidad de Salamanca) tienden a contar también con personal técnico especializado, recursos presupuestarios y planificación estratégica, mientras que las que carecen de políticas se enfrentan a riesgos mayores de pérdida de información digital a largo plazo.

### 3.1.4. *Implementación de estándares y prácticas técnicas*

Uno de los indicadores clave para evaluar la madurez de las estrategias de preservación digital es la adopción de estándares internacionales. Se evaluó el grado de aplicación de cuatro referencias normativas: OAIS, METS, PREMIS y principios FAIR, recogidas en la Tabla 4. Los resultados son los siguientes:

**Tabla 4.** Nivel de adopción de estándares internacionales de preservación digital en repositorios institucionales en España

Estándar o referencia aplicada	n.º de repositorios	%
OAIS	25	50,00
METS	20	40,00
PREMIS	17	34,00
FAIR	16	32,00

Fuente: Elaboración propia.

La implementación de estándares es variable y, en algunos casos, limitada a colecciones específicas o proyectos piloto. Aunque el modelo OAIS es conocido y citado en varias políticas institucionales, su aplicación completa solo se verifica en alrededor de la mitad de los casos. La implementación de PREMIS, por su parte, requiere una estructura de metadatos y herramientas compatibles, lo cual explica su menor adopción.

El cumplimiento de los principios FAIR –esencial para garantizar la reusabilidad de los datos– aún es incipiente en los repositorios institucionales españoles, en parte porque no todos gestionan conjuntos de datos, pero también debido a la falta de formación específica en metadatos de preservación e interoperabilidad semántica.

### 3.1.5. *Prácticas técnicas específicas*

Además de las políticas y estándares, se examinó la aplicación de prácticas técnicas concretamente asociadas a la preservación digital, tales como:

- Migración de formatos: solo 25 repositorios (50 %) indicaron realizarla con regularidad.

- Respaldo distribuido de datos: aplicado por 22 repositorios (44 %).
- Verificación de integridad (*checksums*): detectada en 18 repositorios (36 %).
- Plan de continuidad operativa: implementado en 14 repositorios (28 %).

Estas cifras revelan que más de la mitad de los repositorios no aplican sistemáticamente prácticas básicas de preservación digital, lo que representa una amenaza concreta a la estabilidad de los archivos digitales en el mediano y largo plazo. En muchas cosas, estas tareas quedan relegadas a los servicios informáticos generales de la universidad, sin coordinación directa con los responsables del repositorio.

### 3.1.6. *Financiación específica y recursos humanos*

En cuanto a los recursos disponibles, reflejados en la Tabla 5, se identificó que solo 20 de los 50 repositorios analizados (40 %) cuentan con financiación específica para actividades de preservación digital. El resto depende de presupuestos generales de bibliotecas o unidades de informática, sin partidas diferenciadas.

Asimismo, se constató que apenas 15 repositorios (30 %) disponen de personal técnico con formación específica en preservación digital o metadatos de preservación. Este dato es especialmente relevante, ya que la calidad de las prácticas depende en gran medida del conocimiento y la experiencia de los profesionales a cargo.

Tabla 5. Recursos institucionales disponibles para la preservación digital en repositorios institucionales en España

Recurso disponible	n.º de repositorios	%
Financiación específica	20	40,00
Personal técnico especializado	15	30,00
Plan anual de formación continua	15	30,00

Fuente: Elaboración propia.

Esta situación muestra una fragilidad institucional que afecta la sostenibilidad del sistema. La falta de recursos humanos capacitados y presupuestos estables reduce la capacidad de las universidades para mantener prácticas de conservación digital con estándares adecuados.

### 3.1.7. *Casos de buenas prácticas identificadas*

En el marco del estudio se destacaron algunos casos de repositorios que han implementado estrategias ejemplares en materia de preservación digital:

- GREDOS (Universidad de Salamanca): usa DSpace con integración de METS y PREMIS; tiene política formal de preservación y plan de respaldo.
- Dialnet (Universidad de La Rioja): participación activa en iniciativas de preservación colaborativa, migración periódica de formatos.
- e-Spacio (UNED): política formal, metadatos estructurados, respaldo múltiple y personal capacitado.

Estos casos no solo cumplen con estándares reconocidos, sino que también demuestran un compromiso institucional claro, reflejado en documentación pública, formación de equipos, integración de sistemas y colaboración con redes nacionales e internacionales.

### 3.1.8. *Análisis comparativo por tipología universitaria*

Un análisis adicional permitió identificar algunas diferencias según la tipología y el tamaño de las universidades. Las universidades con mayor volumen de producción científica y con tradición investigadora consolidada (por ejemplo, universidades del G9, la Complutense o la Autónoma de Barcelona) tienden a tener mejores niveles de implementación de estrategias de preservación digital. En cambio, universidades de tamaño medio o reciente creación enfrentan mayores dificultades por restricciones presupuestarias o menor disponibilidad de personal técnico.

Asimismo, se observó que las universidades con estructuras bibliotecarias más integradas y profesionalizadas (como aquellas que participan activamente en REBIUN) tienden a ofrecer políticas de preservación más estructuradas, lo cual subraya la importancia de contar con unidades responsables de planificar, coordinar y evaluar estas tareas de manera continua.

### 3.1.9. *Evolución temporal de las estrategias de preservación*

Una dimensión que merece destacarse es la evolución temporal de las estrategias de preservación en los repositorios reflejadas en la Tabla 6. El análisis de los repositorios con más de 10 años de funcionamiento ( $n = 27$ ) frente a aquellos de creación más reciente ( $n = 23$ ) muestra una correlación entre antigüedad y nivel de desarrollo de políticas de preservación.

**Tabla 6.** Relación entre la antigüedad del repositorio y la existencia de políticas formales de preservación digital

Antigüedad del repositorio	Con política formal	Sin política formal	Total de repositorios
Más de 10 años	10 (37 %)	17(63 %)	27
Menos de 10 años	2 (9 %)	21 (91 %)	23

Fuente: Elaboración propia.

Esto sugiere que la implementación de políticas de preservación digital ha sido progresiva y que aún no ha permeado plenamente en los repositorios más recientes. No obstante, cabe destacar que algunos de estos repositorios jóvenes presentan avances en otras áreas, como el uso de metadatos estructurados o integración con redes temáticas de datos abiertos.

### 3.1.10. *Comparación con iniciativas internacionales*

Para contrastar la situación española, se revisaron informes recientes sobre preservación digital en universidades de Reino Unido, Alemania y Francia. Mientras que en España solo el 24 % de los repositorios tienen políticas detalladas, en Reino Unido esa cifra supera el 70 % (según datos de Jisc, 2022). En Alemania, gracias a la coordinación a través de NFDI (*National Research Data Infrastructure*), la implementación de estándares de preservación como OAIS y PREMIS es mucho más sistemática.

Este contraste evidencia que en España se requiere una estrategia nacional articulada, semejante a los modelos europeos consolidados, que ofrezca directrices y recursos comunes a las universidades para asegurar la conservación de su producción académica y científica.

### 3.1.11. *Implicaciones de la falta de preservación*

Los resultados revelan consecuencias directas de la falta de estrategias de preservación. En al menos 12 repositorios se identificaron documentos obsoletos o con problemas de visualización por haber sido subidos en formatos antiguos (.doc, .wpd, .ps). Algunos repositorios carecían de herramientas para migrar estos documentos a formatos sostenibles (como PDF/A), lo que afecta su accesibilidad.

Asimismo, el 22 % de los repositorios no permiten la exportación masiva de contenidos ni la interoperabilidad con *harvesters* (como OAI-PMH), limitando la circulación del conocimiento y aumentando el riesgo de aislamiento digital.

### 3.1.12. *Papel de las redes interinstitucionales: REBIUN y más allá*

Un elemento positivo destacado en el análisis es el papel que juegan redes como REBIUN y grupos como FESABID en la difusión de buenas prácticas. Las universidades que participan activamente en estas redes muestran mejores indicadores en cuanto a planificación de la preservación digital. En concreto, 18 de los 20 repositorios con políticas publicadas son miembros activos de los grupos técnicos de REBIUN.

Esta correlación evidencia que el intercambio de conocimiento, la estandarización y la cooperación entre instituciones favorecen el desarrollo de estrategias sostenibles. Sin embargo, no todas las universidades están integradas plenamente en estas dinámicas, lo que refuerza la necesidad de políticas públicas de adhesión y fortalecimiento.

3.1.13. *Proyección de sostenibilidad a 10 años*

Finalmente, se elaboró una proyección sobre la sostenibilidad de los repositorios en función de su nivel actual de implementación de estrategias de preservación, reflejada en la Tabla 7. Para ello, se clasificaron los repositorios en tres niveles: alto, medio y bajo, considerando 10 variables técnicas y organizativas. Los resultados fueron:

Tabla 7. Proyección del nivel de sostenibilidad de los repositorios institucionales universitarios a 10 años

Nivel de sostenibilidad proyectada	n.º de repositorios	%
Alto	10	20,00
Medio	18	36,00
Bajo	22	44,00

Fuente: Elaboración propia.

Los repositorios con bajo nivel de sostenibilidad enfrentan riesgos significativos de pérdida de datos, dependencia tecnológica y desactualización normativa. Estos resultados subrayan la urgencia de intervención institucional y nacional para garantizar la preservación del patrimonio científico digital de las universidades.

## 3.2. DISCUSIÓN

Los resultados del presente estudio revelan un panorama complejo y desigual en cuanto a la preservación digital en los repositorios institucionales universitarios en España. Si bien el país ha avanzado notablemente en la consolidación de infraestructuras tecnológicas para el acceso abierto, subsisten importantes debilidades estructurales, organizativas y normativas que comprometen la sostenibilidad de dichos repositorios como garantes del patrimonio científico digital. A partir de las trece dimensiones analizadas, se procede a una discusión detallada de los hallazgos y sus implicaciones para el desarrollo de estrategias institucionales y políticas públicas orientadas a fortalecer la preservación digital en el contexto de la ciencia abierta.

3.2.1. *Infraestructura tecnológica y plataformas utilizadas*

La prevalencia del software DSpace en el 70 % de los repositorios examinados sugiere una estandarización favorable para la interoperabilidad y la adopción de buenas prácticas técnicas. Sin embargo, esta homogeneidad superficial oculta una disparidad significativa en cuanto al grado de personalización, mantenimiento, actualización y explotación de sus funcionalidades avanzadas. Mientras algunas universidades aprovechan plenamente las capacidades de DSpace

para implementar metadatos de preservación y esquemas de control de integridad, otras mantienen configuraciones mínimas que apenas garantizan la operatividad básica del repositorio.

Este hallazgo pone de manifiesto que la elección de una plataforma no es, por sí sola, un indicador de calidad en términos de preservación digital. La sostenibilidad no depende únicamente de la tecnología empleada, sino de la capacidad institucional para gestionarla adecuadamente. En este sentido, la infraestructura técnica debe considerarse un medio y no un fin: lo esencial es la existencia de una planificación institucional explícita que articule los recursos tecnológicos, humanos y normativos con los objetivos de conservación a largo plazo.

### *3.2.2. Tipologías documentales alojadas*

La diversidad de materiales alojados en los repositorios representa un desafío adicional para la preservación. La amplia presencia de artículos científicos y tesis doctorales, frente a la escasa representación de conjuntos de datos de investigación (46 %) y recursos educativos abiertos (32 %), sugiere una concepción limitada del repositorio como mero archivo de publicaciones. Esta visión restringe el potencial del repositorio como espacio integral de gestión de conocimiento y limita su papel en la infraestructura de la ciencia abierta.

Además, la diversidad de formatos y requerimientos técnicos entre los distintos tipos de documentos exige políticas de preservación diferenciadas. Por ejemplo, los conjuntos de datos requieren estrategias de versionado, validación técnica y preservación semántica que van más allá de lo necesario para un artículo en formato PDF. El hecho de que menos de la mitad de los repositorios analizados gestione datos de investigación pone en evidencia una brecha entre las exigencias de la ciencia contemporánea y las capacidades actuales de las universidades para garantizar su conservación.

### *3.2.3. Políticas y estrategias de preservación digital*

Solo el 24 % de los repositorios cuenta con políticas formales y detalladas de preservación digital, lo que plantea un grave problema de gobernanza. Las políticas no solo cumplen una función normativa, sino que también definen responsabilidades, asignan recursos y establecen criterios de evaluación. La ausencia de políticas documentadas implica, por tanto, una falta de visión estratégica institucional y una vulnerabilidad frente a riesgos técnicos, económicos o administrativos.

Este déficit normativo se traduce en prácticas inconsistentes, dependencia de decisiones ad hoc y dificultad para rendir cuentas sobre la gestión del patrimonio digital. Además, la carencia de marcos de referencia internos impide la integración efectiva de los repositorios en otras estrategias universitarias de gestión de la información, como los planes de ciencia abierta, transparencia institucional o innovación docente.

### 3.2.4. *Implementación de estándares y prácticas técnicas*

El análisis muestra que solo entre el 32 % y el 50 % de los repositorios adopta estándares como OAIS, METS, PREMIS o FAIR, siendo esta última especialmente baja. Esta baja implementación limita la interoperabilidad de los repositorios, restringe su integración con redes internacionales y reduce la posibilidad de intercambio, preservación compartida y visibilidad global de los contenidos.

Los estándares internacionales no deben verse como una exigencia técnica externa, sino como una herramienta para asegurar que los contenidos sean comprensibles, accesibles y utilizables a largo plazo. Su adopción implica una transición desde la mera publicación hacia la curaduría digital del conocimiento, lo cual requiere competencias técnicas, planificación institucional y respaldo político. La no adopción de estos estándares deja a los repositorios en un estado de fragilidad y dificulta su evolución hacia modelos sostenibles y colaborativos.

### 3.2.5. *Prácticas técnicas específicas*

Las prácticas concretas de preservación, como la migración de formatos, el respaldo distribuido, la verificación de integridad o la planificación de contingencias, presentan niveles de implementación muy dispares. Solo la mitad de los repositorios realiza migración de formatos y menos del 30 % cuenta con planes de continuidad operativa. Esta situación deja a muchos repositorios expuestos a pérdidas de información, corrupción de archivos, ataques informáticos o fallos tecnológicos.

La ausencia de estas prácticas no responde únicamente a carencias técnicas, sino también a una baja conciencia institucional sobre la importancia de la preservación digital como responsabilidad ética y académica. En este punto, la sensibilización de los equipos directivos y la capacitación del personal técnico son elementos clave para impulsar mejoras sostenibles en las prácticas cotidianas de los repositorios.

### 3.2.6. *Financiación específica y recursos humanos*

Los datos reflejan una escasa disponibilidad de recursos financieros y humanos específicos para la preservación digital. Solo el 40 % de los repositorios cuenta con financiación dedicada y apenas el 30 % dispone de personal técnico especializado. Esta situación contrasta con la complejidad técnica y organizativa que implica la conservación digital a largo plazo y evidencia una disociación entre las expectativas depositadas en los repositorios y los medios efectivamente asignados para su funcionamiento.

La preservación digital no puede sostenerse con estructuras mínimas ni con presupuestos residuales. Requiere una inversión continuada en almacenamiento seguro, formación, desarrollo de herramientas, evaluación y documentación. En este contexto, resulta urgente que las universidades revisen sus modelos de financiación y consideren la preservación como una función esencial y no complementaria de la actividad científica y académica.

### 3.2.7. *Casos de buenas prácticas identificadas*

La identificación de repositorios como GREDOS (Universidad de Salamanca), e-Spacio (UNED) y Dialnet (Universidad de La Rioja) permite afirmar que existen capacidades institucionales reales para implementar buenas prácticas de preservación digital en el sistema universitario español. Estos repositorios no solo cuentan con políticas públicas, sino que también aplican estándares técnicos avanzados, gestionan respaldos distribuidos, controlan la integridad de los objetos digitales y forman a su personal en preservación.

Estos casos confirman que, cuando existe un compromiso institucional respaldado por estructuras formativas y recursos estables, la preservación digital es posible y sostenible. Asimismo, ofrecen modelos replicables para otras universidades, especialmente si se promueve el trabajo colaborativo en red y se crean mecanismos para compartir infraestructura, conocimientos y soluciones tecnológicas.

Es importante destacar que estas buenas prácticas suelen estar asociadas a universidades con trayectorias consolidadas, volúmenes elevados de producción científica y participación activa en redes como REBIUN u OpenAIRE. Esto refuerza la idea de que el contexto institucional influye directamente en la madurez y la calidad de los repositorios y que las políticas públicas deben tener en cuenta estas asimetrías al diseñar estrategias de mejora.

### 3.2.8. *Análisis comparativo por tipología universitaria*

El análisis comparativo evidencia que las universidades más grandes y consolidadas presentan mayores niveles de implementación de estrategias de preservación digital. Esto no es sorprendente, ya que estas instituciones suelen tener mayores capacidades administrativas, técnicas y financieras. Sin embargo, esta constatación también plantea un riesgo: que el desarrollo de la preservación digital se convierta en un fenómeno de élites académicas, dejando atrás a universidades más pequeñas o con menor tradición investigadora.

Frente a ello, resulta fundamental promover políticas de equidad digital que garanticen que todas las universidades públicas, independientemente de su tamaño, puedan preservar su producción científica en condiciones adecuadas. Esto puede lograrse mediante el fomento de plataformas compartidas, la creación de consorcios interuniversitarios, el desarrollo de normativas comunes y la provisión de recursos centralizados desde organismos nacionales.

### 3.2.9. *Evolución temporal de las estrategias de preservación*

La comparación entre repositorios con más de diez años de antigüedad y los de creación más reciente revela que la preservación digital ha sido una preocupación creciente, pero aún no se ha integrado de manera sistemática desde el inicio en los proyectos nuevos. Es decir, muchas universidades han creado repositorios para cumplir con mandatos de acceso abierto, pero no han incorporado desde el comienzo criterios de sostenibilidad a largo plazo.

Este hallazgo invita a revisar los procesos de diseño y planificación de nuevos repositorios institucionales, para que la preservación digital no sea una fase secundaria o reactiva, sino un componente central desde el inicio. Incorporar la preservación desde la concepción de los repositorios permitirá optimizar su arquitectura, formar al personal desde fases tempranas y reducir costes futuros.

### 3.2.10. *Comparación con iniciativas internacionales*

Frente a experiencias internacionales como las de Reino Unido (Jisc), Alemania (NFDI) o Francia (HAL), España presenta una situación más fragmentada y dependiente de iniciativas institucionales individuales. La ausencia de una estrategia nacional de preservación digital limita la coordinación entre universidades, dificulta el desarrollo de estándares comunes y reduce las posibilidades de compartir soluciones tecnológicas y metodológicas.

La creación de un marco nacional permitiría alinear esfuerzos, generar economías de escala y elevar el estándar mínimo de calidad para todos los repositorios. Este marco debería incluir políticas, indicadores de seguimiento, formación, infraestructuras compartidas y recursos financieros. Además, debe articularse con estrategias europeas e internacionales de ciencia abierta para garantizar la interoperabilidad global.

#### 3.2.11. *Implicaciones de la falta de preservación*

El hallazgo de documentos ilegibles o almacenados en formatos obsoletos en algunos repositorios confirma que la falta de políticas y prácticas de preservación digital tiene efectos concretos y perjudiciales. La imposibilidad de acceder a ciertos documentos no solo afecta la visibilidad del conocimiento, sino que también debilita la credibilidad del sistema académico, compromete la reproducibilidad científica y desperdicia recursos públicos invertidos en investigación.

Desde una perspectiva ética, la preservación digital es un deber de las universidades hacia la sociedad. El conocimiento financiado con fondos públicos debe permanecer accesible para las futuras generaciones. No garantizar su conservación implica una forma de pérdida cultural, patrimonial y científica que puede ser irreparable.

#### 3.2.12. *Papel de las redes interinstitucionales: REBIUN y más allá*

El papel de REBIUN y otras redes colaborativas ha sido clave en la promoción de buenas prácticas y en la generación de documentos orientadores. Sin embargo, la participación activa en estas redes no es uniforme entre todas las universidades, lo cual genera desigualdades. Fortalecer estas redes y convertirlas en actores estratégicos del sistema nacional de preservación digital es un paso necesario para superar la fragmentación actual.

Se propone, por tanto, que las redes universitarias asuman un rol de liderazgo coordinado, funcionando como agencias de certificación, formación, asesoría técnica y seguimiento de estándares. Asimismo, pueden ser el canal para articularse con redes europeas como COAR (*Confederation of Open Access Repositories*), OpenAIRE o SPARC Europe, ampliando las posibilidades de cooperación internacional.

### 3.2.13. *Proyección de sostenibilidad a 10 años*

La proyección realizada en este estudio, que sitúa a un 44 % de los repositorios en una situación de baja sostenibilidad, debe ser interpretada como una alerta para los tomadores de decisión. Si no se invierte de forma inmediata en reforzar los elementos técnicos y organizativos de los repositorios, una parte importante del conocimiento almacenado corre el riesgo de volverse inaccesible o degradarse.

La sostenibilidad no puede depender únicamente de la voluntad o capacidad individual de cada universidad. Debe ser abordada como una responsabilidad colectiva del sistema universitario y del Estado. Esto requiere una planificación nacional que incluya financiación estable, incentivos para la colaboración, indicadores de calidad y mecanismos de rendición de cuentas.

## 4. CONCLUSIONES

Los hallazgos obtenidos en el presente estudio permiten confirmar la hipótesis planteada en la introducción: a pesar de contar con una infraestructura tecnológica ampliamente consolidada y con un ecosistema de repositorios institucionales universitarios extendido a lo largo del sistema de educación superior en España, persisten notables carencias estructurales, organizativas, normativas y formativas en la implementación efectiva de estrategias de preservación digital. El análisis detallado de cincuenta repositorios activos en universidades públicas ha permitido no solo identificar patrones de avance, sino también revelar lagunas significativas en materia de sostenibilidad, estándares, planificación y formación técnica.

En primer lugar, se constató que, si bien la plataforma DSpace es utilizada por el 70 % de los repositorios, la mera adopción de una tecnología común no garantiza por sí sola buenas prácticas de preservación. La gestión eficaz de un repositorio depende de la existencia de políticas institucionales, competencias técnicas, asignación de recursos y compromiso estratégico. Esta observación confirma la necesidad de trascender el enfoque puramente instrumental para situar la preservación digital en el centro de la política universitaria de gestión del conocimiento.

En cuanto a la tipología de contenidos, la mayoría de los repositorios se concentra en la conservación de artículos científicos y tesis doctorales, mientras que los conjuntos de datos de investigación y los recursos educativos abiertos están escasamente representados. Esta limitación revela una concepción aún restringida del acceso abierto, centrada más en la publicación formal que en la gestión integral del ciclo de vida de la producción académica. El

bajo porcentaje de repositorios que gestionan datos de investigación (46 %) obstaculiza la implementación efectiva de los principios FAIR y restringe el potencial de reutilización de la ciencia en entornos interdisciplinarios.

Otro de los aspectos centrales del estudio ha sido el análisis de las políticas institucionales de preservación digital. La investigación ha mostrado que solo el 24 % de los repositorios analizados dispone de una política formal y detallada sobre esta materia. Esta ausencia de marcos normativos compromete seriamente la continuidad y la calidad del servicio, ya que impide una gestión sistemática y planificada de los procesos de conservación digital. La falta de políticas, además, genera dependencia de decisiones individuales o de condiciones presupuestarias cambiantes, dejando a los repositorios expuestos a riesgos estructurales.

En cuanto a la adopción de estándares internacionales como OAIS, METS, PREMIS y FAIR, se ha comprobado que su nivel de implementación es aún limitado, oscilando entre el 32 % y el 50 % según el caso. Esta carencia afecta la interoperabilidad de los repositorios, su integración con redes internacionales de datos abiertos y la posibilidad de garantizar una trazabilidad adecuada de los objetos digitales a lo largo del tiempo. La preservación digital, para ser efectiva, no puede prescindir de estos marcos de referencia que permiten documentar, verificar y recuperar los contenidos en condiciones de fiabilidad y autenticidad.

Asimismo, las prácticas técnicas específicas de preservación presentan un desarrollo desigual. Mientras que algunas instituciones han incorporado procedimientos de migración de formatos, respaldos distribuidos, controles de integridad y planes de continuidad operativa, más del 50 % de los repositorios carecen de estos mecanismos o los aplican de manera parcial o informal. Esta situación refleja tanto una falta de formación técnica como una escasa priorización institucional del problema.

La dimensión financiera y humana constituye otro de los pilares débiles del sistema actual. Solo el 40 % de los repositorios cuenta con financiación específica para tareas de preservación digital, y apenas el 30 % dispone de personal con formación especializada en la materia. Esta precariedad estructural limita la capacidad de las universidades para garantizar la sostenibilidad de los contenidos almacenados y obstaculiza la actualización continua de sus infraestructuras. Además, la falta de inversión en formación impide el desarrollo de una cultura institucional sólida en torno a la preservación.

No obstante, el estudio también ha permitido identificar buenas prácticas institucionales que pueden servir como referencia para otras universidades. Repositorios como GREDOS (Universidad de Salamanca), Dialnet (Universidad de La Rioja) y e-Spacio (UNED) han desarrollado políticas sólidas, implementado estándares reconocidos y formando equipos técnicos capaces de garantizar la calidad y la sostenibilidad del servicio. Estos casos demuestran que es posible articular prácticas eficaces de preservación digital en el ámbito universitario público, siempre que exista voluntad institucional, liderazgo técnico y asignación de recursos adecuados.

En el análisis comparativo, se evidenció también una brecha por tamaño institucional y trayectoria. Las universidades más grandes o con mayor tradición investigadora han desarrollado con mayor éxito políticas de preservación, mientras que las de reciente creación o de menor tamaño presentan más dificultades. Esta desigualdad estructural plantea la necesidad

de implementar estrategias de cooperación interuniversitaria, consorcios regionales o servicios compartidos que garanticen un nivel mínimo de calidad y sostenibilidad para todos los repositorios del sistema.

La dimensión temporal también ha sido relevante. Aunque se observa una tendencia positiva en el desarrollo de políticas en los repositorios más antiguos, esta evolución aún no se traduce de forma sistemática en las nuevas plataformas creadas en la última década. El hecho de que muchos de los repositorios más recientes no integren la preservación digital desde su fase de diseño refuerza la urgencia de revisar los protocolos de planificación y exigir, desde el inicio, el cumplimiento de requisitos técnicos y normativos de conservación.

Comparado con países del entorno europeo como Reino Unido, Alemania o Francia, el sistema español de preservación digital presenta debilidades evidentes. La falta de una estrategia nacional unificada, la dispersión de esfuerzos y la escasa coordinación entre instituciones dificultan el desarrollo de soluciones conjuntas. Frente a ello, el presente estudio sostiene que es urgente promover una política pública estatal de preservación digital universitaria, articulada desde el Ministerio de Ciencia, las agencias nacionales de evaluación y redes como REBIUN, en coordinación con estándares europeos.

La ausencia de preservación digital no es una amenaza abstracta, sino una realidad comprobable: documentos ilegibles, formatos obsoletos y barreras de acceso ya afectan a varios repositorios españoles. Estos problemas tienen implicaciones concretas en términos de reproducibilidad científica, visibilidad académica, pérdida de memoria institucional y desaprovechamiento de la inversión pública en investigación. La preservación digital, por tanto, no puede ser vista como un accesorio técnico, sino como un pilar fundamental de la infraestructura del conocimiento.

En este marco, las redes interuniversitarias como REBIUN han demostrado ser actores fundamentales en la promoción de estándares, la formación de profesionales y la coordinación técnica entre instituciones. Su rol debe ser fortalecido mediante mecanismos de financiación estable, reconocimiento institucional y colaboración internacional.

Finalmente, la proyección de sostenibilidad a diez años, elaborada a partir de los datos obtenidos, muestra que el 44 % de los repositorios se encuentra en una situación de baja sostenibilidad. Esta cifra representa una alerta para las autoridades académicas y políticas, que deben actuar con urgencia para revertir esta tendencia y garantizar la preservación del patrimonio científico del país.

## REFERENCIAS BIBLIOGRÁFICAS

- Abad García, M. F. *et al.* (2022). Les universitats espanyoles i la ciència oberta: un estudi sobre barreres i elements afavoridors. *BiD: textos universitaris de biblioteconomia i documentació*, 49. <https://bid.ub.edu/en/49/abad.htm>
- Abadal, E. *et al.* (2019). El plan S para acelerar el acceso abierto: contexto, retos y debate generado. *Hipertext.net*, 19, 75-83. <https://doi.org/10.31009/hipertext.net.2019.i19.06>

- Azorín, C. *et al.* (2020). *Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación*. Colección Estudios e Informes. REBIUN. <https://doi.org/10.20350/digitalCSIC/15055>
- Barba-Salazar, J. *et al.* (2024). Ventajas y desventajas de los repositorios digitales en instituciones de Educación Superior. *Digital Publisher CEIT*, 8(4), 371-382. <https://doi.org/10.33386/593dp.2024.4.2522>
- Batthyány, K. *et al.* (2023). *Iniciativas y regulaciones multinivel para la ciencia abierta*. Documentos de trabajo, 91. Fundación Carolina y Folec-Clacso. [https://www.fundacioncarolina.es/wp-content/uploads/2023/11/DTFC91\\_web.pdf](https://www.fundacioncarolina.es/wp-content/uploads/2023/11/DTFC91_web.pdf)
- Castro Leal, L. (2024). Retos y oportunidades en la preservación digital del patrimonio documental. La experiencia de Repsol. *Revista Española de Documentación Científica*, 47(2), e384. <https://doi.org/10.3989/redc.2023.2.1511>
- De-Filippo, D. *et al.* (2023). Mapping open science at Spanish universities. Analysis of higher education systems. *El profesional de la Información*, 32(4). <https://doi.org/10.3145/epi.2023.jul.06>
- De-Giusti, M. (2023). Talleres de preservación digital en repositorios institucionales. Preservación Digital de repositorios. En *Conferencia Internacional BIREDIAL-ISTEE*. <https://sedici.unlp.edu.ar/bitstream/handle/10915/160116/Presentaci%C3%B3n.pdf-PDFA.pdf?sequence=1&isAllowed=y>
- Ferreras Torres, P. (2010). *Preservación digital en repositorios institucionales: GREDOS* (Trabajo de Fin de Máster, Universidad de Salamanca). <http://hdl.handle.net/10366/83130>
- Hallo Carrasco, M. y De la Fuente, P. (2010). Preservación digital en el contexto de repositorios institucionales universitarios. *Revista Cubana de Ciencias Informáticas*, 7(2), 3-18. <https://www.redalyc.org/pdf/3783/378343670001.pdf>
- Library of Congress. (2008). *Diccionario de datos PREMIS: de metadatos de Preservación versión 2.0*. [https://www.loc.gov/standards/premis/PREMIS\\_es.pdf](https://www.loc.gov/standards/premis/PREMIS_es.pdf)
- Monteagudo-Haro, P. y Prieto-Gutiérrez, J. J. (2024). Open Research Data in Spanish University Repositories. *Revista Española de Documentación Científica*, 47(3). <https://doi.org/10.3989/redc.2024.3.1581>
- Ochoa-Gutiérrez, J. *et al.* (2021). Experiencias de gestión de los procesos de preservación digital a partir del modelo OAIS en repositorios institucionales. *Anales de Documentación*, 24(1), 1-18. <https://doi.org/10.6018/analesdoc.428141>
- Raventós Pajares, P. (2017). Metadatos de preservación: el modelo PREMIS y otros. *Boletín de la ANA-BAD*, 65(4), 108-151. <https://dialnet.unirioja.es/servlet/articulo?codigo=6714238>
- Segundo Martín, M. L. *et al.* (2022). *Preservación digital de las tesis en los repositorios de REBIUN*. Colección Estudios e Informes. REBIUN. <http://hdl.handle.net/20.500.11967/1171>
- Suárez, J. M. (2024). *Preservación digital: un estudio de adecuación sobre PREMIS* (Trabajo Fin de Postgrado, Universidad Nacional de La Plata). <https://sedici.unlp.edu.ar/handle/10915/169707>


## LOS SERVICIOS BIBLIOTECARIOS DE APOYO A LA INVESTIGACIÓN EN LA UNIVERSIDAD DE SALAMANCA

### *RESEARCH SUPPORT LIBRARY SERVICES AT THE UNIVERSITY OF SALAMANCA*

Almudena MANGAS-VEGA<sup>1</sup> y Sonia SANTIAGO ROMÁN<sup>2</sup>

<sup>1</sup> *Universidad de Salamanca. España*

[almumvega@usal.es](mailto:almumvega@usal.es)

 <https://orcid.org/0000-0002-3464-3624>

<sup>2</sup> *Universidad de Salamanca. España*

[sonias@usal.es](mailto:sonias@usal.es)

**RESUMEN:** En el actual contexto de transformación digital y ciencia abierta, los servicios bibliotecarios universitarios han ampliado sus funciones tradicionales para ofrecer un apoyo estratégico en la gestión de los datos de investigación. En el caso del Servicio de Bibliotecas de la Universidad de Salamanca (USAL), este apoyo se concreta en la orientación a investigadores sobre el ciclo de vida de los datos, desde su planificación hasta su difusión y preservación, alineándose con los principios FAIR y con los requisitos de convocatorias nacionales e internacionales. El presente trabajo identifica y analiza los servicios más relevantes vinculados a datos de investigación que se han implementado en la USAL en los últimos años, atendiendo a criterios de innovación, demanda por parte de la comunidad académica y necesidad de nuevas competencias técnicas y formativas. Entre ellos destacan la formación en gestión de datos, el asesoramiento en planes de gestión exigidos por financiadores, la preservación en el repositorio institucional y la integración con iniciativas de ciencia abierta. Asimismo, se describen los principales retos detectados, como la necesidad de concienciación y alfabetización en datos, la heterogeneidad disciplinar en la gestión de datos y las limitaciones estructurales y de personal especializado. Estos desafíos evidencian la urgencia de consolidar equipos bibliotecarios capacitados para acompañar a los investigadores en entornos cada vez más complejos. El estudio del caso de los servicios bibliotecarios de apoyo a la investigación de la USAL muestra cómo los servicios bibliotecarios pueden contribuir activamente a la calidad, la visibilidad y el cumplimiento normativo de la investigación a través de una gestión eficaz de los datos científicos.

**PALABRAS CLAVE:** gestión de datos de investigación; bibliotecas académicas; ciencia abierta; principios FAIR; alfabetización en datos; Universidad de Salamanca.

**ABSTRACT:** In the current context of digital transformation and open science, university library services have expanded their traditional roles to provide strategic support in research data management. At the University of Salamanca Library Service (USAL), this support is reflected in guiding researchers through the entire research data lifecycle – from planning to dissemination and preservation – aligning with FAIR principles and the requirements of national and international funding calls. This study identifies and analyzes the most relevant research data-related services implemented at USAL in recent years, based on criteria such as innovation, demand from the academic community, and the need for new technical and training competencies. Key services include training in data management, advising on data management plans required by funders, long-term preservation in the institutional repository, and integration with open science initiatives. The paper also describes major challenges, such as raising awareness and promoting data literacy, disciplinary diversity in data practices, and structural limitations including staffing and expertise. These challenges highlight the urgent need to consolidate skilled library teams capable of supporting researchers in increasingly complex environments. This case study demonstrates how library services can actively contribute to the quality, visibility, and regulatory compliance of research through effective research data management.

**KEYWORDS:** research data management; academic libraries; open science; FAIR principles; data literacy; University of Salamanca

## 1. INTRODUCCIÓN

En el contexto actual de transformación digital y ciencia abierta, las bibliotecas universitarias han ampliado sus funciones tradicionales para ofrecer un apoyo estratégico en la gestión de los datos de investigación (Dobrevá *et al.*, 2010). El V Plan Estratégico de la Red Española de Bibliotecas Universitarias de 2024 a 2027 (REBIUN, 2023) destaca la importancia de reformular el modelo de biblioteca para adaptarse a las novedades de la transformación digital y promover la ciencia abierta, reconociendo el papel esencial de las bibliotecas como agentes estratégicos en este nuevo ecosistema. Tal como plantea Agosti (2018), las bibliotecas universitarias deben asumir un rol activo en la orientación y el acompañamiento a los investigadores, incorporando servicios especializados que respondan a las nuevas exigencias de la ciencia abierta.

En línea con estas directrices, el Servicio de Bibliotecas de la Universidad de Salamanca (USAL) ha desarrollado una serie de servicios orientados a facilitar la gestión de los datos de investigación, desde su planificación hasta su difusión y preservación.

Estos servicios incluyen la alfabetización y la formación en gestión de datos, el asesoramiento en planes de gestión exigidos por financiadores, la preservación en el repositorio institucional, la integración con iniciativas de ciencia abierta y la recogida de métricas de lectura y uso.

Además, el Servicio de Bibliotecas de la USAL participa activamente en la formación continua de la comunidad universitaria, colaborando con el Instituto Universitario de Ciencias

de la Educación (IUCE) y con los centros de la Universidad en la programación para el Personal Docente e Investigador (PDI), en la formación transversal de la Escuela de Doctorado y en la formación específica para el Personal Técnico de Gestión, Administración y Servicios (PTGAS). La formación en gestión de datos de investigación se ha integrado como un componente esencial en estos programas, reconociendo la necesidad de dotar a los investigadores y al personal de apoyo de las competencias necesarias para manejar los datos de manera efectiva. Se han desarrollado módulos formativos específicos sobre la importancia de los metadatos en la gestión de datos, las herramientas disponibles para efectuar esa gestión y las buenas prácticas en la preservación y el acceso a los datos de investigación (Johnston, 2014). Estos programas formativos también incluyen talleres prácticos sobre la elaboración de planes de gestión de datos (DMP), la selección de repositorios de datos adecuados y la aplicación de los principios FAIR (Findable, Accessible, Interoperable, Reusable) a los datos de investigación (Jones, 2011; Voutssás Márquez, 2012).

En el ámbito de la evaluación de la investigación, desde los servicios de apoyo a la investigación del Servicio de Bibliotecas de la USAL se colabora con grupos de asesoría, con unidades de investigación, con investigadores e incluso con agentes de evaluación, asesorando en convocatorias competitivas y en la búsqueda y buen uso de criterios y métricas. La creciente importancia de los datos de investigación en la evaluación de la calidad y el impacto de la investigación ha llevado al Servicio de Bibliotecas a desarrollar servicios especializados en el apoyo a los investigadores en la gestión y el uso de métricas relacionadas con los datos. Estos servicios incluyen la recopilación y el análisis de datos de uso, lectura y citación de conjuntos de datos o la elaboración de informes sobre el impacto de los datos de investigación en la comunidad científica y en la sociedad en general.

Asimismo, el proyecto BibliODS del Servicio de Bibliotecas de la USAL facilita la vinculación de la producción científica con los Objetivos de Desarrollo Sostenible (ODS), permitiendo una evaluación precisa del impacto de las investigaciones y generando informes fiables que eviten interpretaciones parciales o sesgadas sobre la contribución real de los investigadores a los ODS. La gestión adecuada de los datos de investigación es fundamental para garantizar la transparencia y la reproducibilidad de los resultados, lo que a su vez fortalece la credibilidad de la investigación y facilita su contribución a los ODS (Corti *et al.*, 2014). Estos informes permiten a los investigadores y a las instituciones comprender y comunicar el impacto de su trabajo en el desarrollo sostenible.

En este contexto, los servicios bibliotecarios de apoyo a la investigación se consolidan como aliados fundamentales de la institución y de los investigadores, favoreciendo que el trabajo científico sea más eficiente, colaborativo y accesible, y repercutiendo en una mayor rentabilidad en recursos, tiempo y esfuerzos en programas y convocatorias institucionales, nacionales e internacionales. La gestión eficaz de los datos de investigación no solo optimiza el proceso de investigación, sino que también facilita la colaboración entre investigadores y la reutilización de los datos, lo que a su vez impulsa la innovación y el avance del conocimiento, todo ello dentro del marco de la ciencia abierta. Se han desarrollado servicios de apoyo que facilitan la colaboración entre investigadores en la gestión de datos, incluyendo herramientas para compartir y anotar datos o funcionalidades en plataformas para la gestión colaborativa de

proyectos de investigación y servicios de consultoría (individual o por unidad) sobre buenas prácticas en la gestión de datos.

En la última convocatoria de sexenios de investigación publicada por la Agencia Nacional de Evaluación y Acreditación Española, ANECA (BOE, 2024, p. 305), se reconoce explícitamente que las bibliotecas universitarias y científicas, así como las unidades de apoyo a la investigación, pueden «facilitar asesoramiento a las personas solicitantes en relación con las dimensiones, métricas y fuentes pertinentes y confiables en cada caso». Este reconocimiento oficial consolida el papel estratégico de las bibliotecas en los procesos de evaluación de la actividad investigadora, especialmente en lo que respecta a la gestión y el uso de datos de investigación, una de las nuevas tipologías (como aportaciones susceptibles de ser evaluadas) incluidas en esta convocatoria recientemente.

La ANECA se adhiere a DORA (San Francisco Declaration on Research Assessment) y a COARA (Coalition for Advancing Research Assessment) e impulsa, así, un cambio en la evaluación de la investigación.

La adhesión a DORA supone la evaluación de los artículos científicos y no solo de las revistas en que se publican y la consideración y el impacto de todos los resultados de investigación incluidos datos y software.

El compromiso con COARA aboga por evaluar prácticas, criterios y herramientas basados en evidencias sólidas y hacer que los datos estén disponibles abiertamente para la recopilación de evidencia e investigación. La ANECA reconoce, de esta forma, la importancia de los datos de investigación como una forma válida de contribución científica y establece criterios y directrices para su evaluación, lo que ha impulsado a las bibliotecas universitarias a desarrollar servicios de apoyo especializados en este ámbito.

Además, en determinadas áreas científicas, ANECA admite como aportaciones válidas los conjuntos de datos de investigación, siempre que se trate de datos FAIR, que cumplan con los requisitos de acceso abierto, estén depositados en repositorios reconocidos y vayan acompañados de la correspondiente documentación técnica y de impacto. Esta inclusión de los datos de investigación como una forma válida de contribución científica subraya la creciente importancia de la gestión de datos en la investigación y la necesidad de servicios bibliotecarios que apoyen a los investigadores en esta área. Por ello, se han desarrollado servicios de apoyo que facilitan el depósito de datos en repositorios de datos, incluyendo la preparación de metadatos, la selección de licencias de datos y la generación de identificadores persistentes para los datos.

Este cambio normativo responde a una evolución en los modelos de evaluación y refuerza la necesidad de contar con servicios especializados que acompañen al investigador en el diseño, la gestión, la documentación y la visibilidad de sus datos. La evaluación de la investigación está evolucionando hacia un enfoque más amplio que considera no solo las publicaciones, sino también los datos, el software, los métodos y otros resultados de la investigación. Las bibliotecas universitarias están desempeñando un papel cada vez más importante en la promoción de este cambio, ofreciendo servicios de formación, asesoramiento y apoyo en la gestión de datos. En muchos casos se han desarrollado servicios de formación y asesoramiento que ayudan a los investigadores a comprender y aplicar los nuevos modelos de evaluación, incluyendo talleres

sobre la gestión de datos de investigación, guías sobre la evaluación de datos y servicios de consultoría sobre la elaboración de portafolios de datos.

En este contexto, todos los servicios descritos —desde la formación en competencias digitales y la elaboración de planes de gestión de datos hasta el depósito en el repositorio institucional, el asesoramiento en convocatorias competitivas o el desarrollo de herramientas como BibliODS— configuran un ecosistema de apoyo integral que facilita el cumplimiento de los principios de la ciencia abierta. La gestión adecuada de los datos de investigación es un componente esencial de este nuevo paradigma, ya que facilita el acceso, la reutilización y la transparencia de los resultados de la investigación. También ha sido grande el desarrollo de servicios que promueven la ciencia abierta, incluyendo plataformas para compartir datos, herramientas para su publicación y servicios de apoyo para la gestión de datos en proyectos de ciencia abierta.

Estos servicios no solo optimizan la gestión de los datos de investigación, sino que también ayudan a visibilizar el valor del trabajo científico y a generar evidencias concretas y trazables para los procesos de evaluación (Cox, Pinfield y Smith, 2019). La disponibilidad y la calidad de los datos de investigación son fundamentales para garantizar la reproducibilidad de los resultados y para facilitar la evaluación del impacto de la investigación. La USAL ha desarrollado servicios de apoyo que ayudan a los investigadores a documentar y contextualizar sus datos, lo que facilita su comprensión y reutilización por parte de otros investigadores y de la sociedad en general.

Así, las bibliotecas universitarias, con equipos especializados y comprometidos, se convierten en actores clave para la calidad, la transparencia y el impacto de la investigación en el marco actual. El presente estudio de caso tiene como objetivo analizar en profundidad los servicios de apoyo a la investigación del Servicio de Bibliotecas de la Universidad de Salamanca, con un enfoque especial en la gestión de datos de investigación, la formación relacionada y sus métricas de investigación. Se busca comprender cómo estos servicios contribuyen al desarrollo de la ciencia abierta y al fortalecimiento de la investigación en la institución, con un énfasis particular en cómo apoyan la gestión, la formación y la defensa de evidencias de calidad e impacto relacionadas con los datos de investigación. Se analizarán en detalle los servicios de formación en gestión de datos, asesoramiento en planes de gestión de datos, apoyo en metadatos, apoyo en métricas de datos y gestión del repositorio institucional, identificando las buenas prácticas, los desafíos y las oportunidades de mejora en cada uno de ellos.

## 2. METODOLOGÍA

Para este estudio, se adoptó una metodología cualitativa basada en el análisis de caso, centrada en los servicios de apoyo a la investigación del Servicio de Bibliotecas de la Universidad de Salamanca (USAL), con especial énfasis en la gestión de datos de investigación (RDM: Research Data Management). La selección de este enfoque responde a la necesidad de comprender en profundidad las prácticas, los desafíos y las oportunidades que enfrentan las bibliotecas universitarias en el contexto de la ciencia abierta y la transformación digital.

La primera fase consistió en una revisión exhaustiva de los servicios y las actividades cotidianas del área de Apoyo a la Investigación de la USAL. Se establecieron criterios específicos para la selección de los elementos a analizar: relevancia para la comunidad académica, cambios significativos en los últimos años, desafíos en su planificación y ejecución y la necesidad de nuevas competencias por parte del personal bibliotecario. Esta revisión permitió identificar prácticas clave y áreas de mejora en la prestación de servicios relacionados con el RDM.

Se prestó especial atención a los servicios de formación tanto presenciales como telemáticos, en gestión de datos, asesoramiento en planes de gestión y apoyo en métricas de datos, reconociendo su importancia para el éxito de la investigación en la era digital (Tenopir *et al.*, 2012). Se analizaron en detalle los contenidos y la metodología de los programas de formación en gestión de datos, la eficacia de los servicios de asesoramiento en la elaboración de planes de gestión de datos y la utilidad de las herramientas y los servicios de apoyo en métricas de datos para los investigadores. Para autores como Cárdenas (2021), la formación en gestión de datos debe estar estrechamente alineada con las necesidades reales de los investigadores y con el ciclo de vida de los datos, lo que implica un diseño flexible y adaptable de los programas formativos.

Posteriormente, se realizó un análisis detallado de los elementos seleccionados, describiendo sus características, objetivos y resultados esperados. Este análisis se complementó con la identificación y la descripción de los obstáculos que afectan el desempeño efectivo de estos servicios, tales como limitaciones de recursos, barreras institucionales y la necesidad de formación especializada. Se examinaron en profundidad los desafíos relacionados con la concienciación y la alfabetización en datos, la heterogeneidad disciplinar en la gestión de datos y la necesidad de desarrollar métricas de datos que reflejen adecuadamente el impacto de los datos de investigación. Se analizaron las estrategias implementadas por la USAL para superar estos obstáculos, incluyendo la colaboración con otras unidades de la universidad, la adaptación de los servicios a las necesidades de los usuarios, la promoción de la ciencia abierta y la evaluación continua de los servicios.

La metodología empleada se alinea con enfoques utilizados en estudios similares. Por ejemplo, González-Solar (2018) destaca la importancia de los estudios de usuarios en el diseño de servicios bibliotecarios de apoyo a la investigación, utilizando técnicas como la observación, las estadísticas, las entrevistas y las encuestas para comprender las necesidades de los investigadores.

Asimismo, Camón, Rey y Balagué (2012) también la utilizaron como paso previo a cuestionarios web dirigidos a la dirección de bibliotecas universitarias para evaluar los servicios de apoyo a la investigación en Cataluña, subrayando la necesidad de adaptar estos servicios a las particularidades de cada institución. Estos estudios resaltan la importancia de comprender las necesidades de los investigadores en relación con la gestión de datos, la formación y las métricas de datos para poder ofrecer servicios de apoyo efectivos. Se analizaron las diferentes metodologías y técnicas utilizadas en estos estudios, evaluando su aplicabilidad al contexto de la USAL.

Además, se consideraron las recomendaciones de la Biblioteca Nacional de España (2020) sobre la necesidad de estrategias y normas alineadas para la gestión de datos de investigación,

destacando el papel activo de las bibliotecas en el desarrollo de nuevos servicios y herramientas. Estas recomendaciones subrayan la necesidad de que las bibliotecas universitarias desarrollen políticas y estrategias claras para la gestión de datos de investigación, así como servicios y herramientas que apoyen a los investigadores en la implementación de estas políticas y estrategias. Se analizaron las recomendaciones de la Biblioteca Nacional de España en relación con la formación en gestión de datos, el asesoramiento en planes de gestión de datos y el apoyo en métricas de datos, evaluando su relevancia para el contexto de la USAL.

La metodología aplicada en este estudio, pues, combina la revisión y el análisis de servicios existentes con la identificación de obstáculos y la consideración de enfoques metodológicos utilizados en estudios similares. Este enfoque integral permite una comprensión profunda de las prácticas de apoyo a la investigación en el Servicio de Bibliotecas de la USAL, ofreciendo *insights* valiosos para la mejora continua de los servicios bibliotecarios en el contexto de la ciencia abierta y la gestión de datos de investigación. Se busca obtener una visión completa de cómo la biblioteca apoya a los investigadores en la gestión, la formación y las métricas relacionadas con los datos de investigación, identificando las fortalezas y las debilidades de los servicios existentes y proponiendo recomendaciones para su mejora. Se analizarán las diferentes perspectivas y enfoques sobre la gestión de datos de investigación, la formación y las métricas de datos, buscando identificar las mejores prácticas y las áreas donde se necesita mayor desarrollo.

### 3. RESULTADOS Y DISCUSIÓN

El siguiente apartado presenta los principales resultados del análisis realizado sobre los servicios de apoyo a la investigación del Servicio de Bibliotecas de la USAL, estructurándolos en cuatro puntos clave: el marco estratégico nacional; el ecosistema de la ciencia abierta y su impacto en las prácticas profesionales; la formación en principios FAIR y tipologías de datos, y los servicios de apoyo a la investigación propiamente dichos, que incluyen los planes de gestión de datos, los repositorios institucionales, los metadatos y el acompañamiento en convocatorias y evaluaciones.

#### 3.1. MARCO ESTRATÉGICO

El V Plan Estratégico de REBIUN (2024-2027) establece un marco de referencia clave para las bibliotecas universitarias y científicas españolas. Este plan reconoce a las bibliotecas como agentes estratégicos en la transformación digital y en el impulso de la ciencia abierta.

Entre sus líneas estratégicas destacan:

- La biblioteca en el contexto de la transformación digital, con nuevos servicios de apoyo a la docencia, la investigación y la transferencia del conocimiento.
- La biblioteca como agente estratégico para la ciencia abierta, promoviendo políticas institucionales, buenas prácticas y sostenibilidad en los modelos de comunicación científica.

En este marco, el Servicio de Bibliotecas de la Universidad de Salamanca (USAL) ha alineado sus políticas bibliotecarias con las directrices de REBIUN, desarrollando estrategias específicas para fortalecer los servicios de apoyo a la investigación. Esto incluye la creación de infraestructuras digitales; la formación del personal bibliotecario en competencias clave, y la integración de servicios de asesoramiento en gestión de datos, métricas y publicación científica. Todo esto se ha implementado dentro de marcos estratégicos institucionales (Universidad de Salamanca, 2018, 2023). Se ha reconocido la importancia de la gestión de datos de investigación como un componente fundamental de la ciencia abierta y se han desarrollado servicios específicos para apoyar a los investigadores en esta área, implementando procedimientos específicos para la gestión de datos de investigación, incluyendo políticas de acceso abierto a los datos, directrices para la elaboración de planes de gestión de datos y recomendaciones para la selección de repositorios de datos. Se han establecido alianzas con otras instituciones y organizaciones para promover la gestión de datos de investigación y compartir buenas prácticas.

Además, los recientes cambios en las convocatorias de ANECA han reforzado la necesidad de estos servicios. Desde 2024, se exige que las aportaciones a los sexenios estén depositadas en repositorios de acceso abierto, incluyendo identificadores persistentes (DOI o handle) y metadatos completos. También se valora la inclusión de *datasets* FAIR y se promueve la evaluación cualitativa de la producción científica, reconociendo aportaciones no tradicionales como software, catálogos o audiovisuales. Esto ha elevado el papel de las bibliotecas como mediadoras entre investigadores y criterios de evaluación, consolidando su función estratégica. ANECA ha reconocido la importancia de los datos de investigación como un resultado valioso de la investigación y ha establecido criterios específicos para su evaluación. Se valora especialmente la calidad de los metadatos de los datos, la accesibilidad de los datos y la reutilización potencial de los datos.

La política de ciencia abierta de la Universidad de Salamanca, en la que el Servicio de Bibliotecas ha tenido un papel fundamental, se publicó en abril de 2025 y se desarrolla en varias dimensiones entre las que se encuentran los datos abiertos de investigación, importantes para poder verificar los resultados de la investigación ya que permiten la reproducibilidad de la misma (Universidad de Salamanca, 2025). La forma de gestionar y compartir los datos de la investigación depende del tipo de datos y de la cultura de las distintas disciplinas y ámbitos científicos. A efectos de esta política, los datos de investigación son cualquier información digital o analógica que haya sido recogida, creada o reutilizada para producir o validar los resultados de una investigación.

### 3.2. ECOSISTEMA DE LA CIENCIA ABIERTA Y CAMBIO EN LAS PRÁCTICAS PROFESIONALES

El ecosistema de la ciencia abierta representa un cambio de paradigma en la forma en que se produce, comparte y evalúa el conocimiento científico. Este modelo se basa en la apertura, la transparencia y la colaboración, y requiere una transformación profunda tanto en las prácticas de los investigadores como en los servicios de apoyo que ofrecen las bibliotecas universitarias.

En este nuevo entorno, las infraestructuras digitales, el *big data*, los servicios en la nube y las herramientas de análisis de datos desempeñan un papel fundamental. Los investigadores necesitan adquirir nuevas competencias para gestionar eficazmente sus datos, utilizar repositorios, aplicar licencias abiertas y participar en redes de colaboración. La ciencia abierta promueve la reutilización de los datos de investigación, lo que requiere que los investigadores documenten y describan sus datos de manera clara y completa. Las bibliotecas universitarias están desempeñando un papel clave en la formación de los investigadores en estas nuevas competencias.

Las bibliotecas universitarias, en colaboración con las unidades de gestión de la investigación, desempeñan un rol clave en el acompañamiento de los investigadores en este nuevo escenario. Las bibliotecas están ofreciendo servicios de asesoramiento sobre la gestión de datos, la selección de repositorios y la aplicación de licencias de datos. También están facilitando el acceso a herramientas y plataformas para la gestión y el análisis de datos. En el caso de ANECA, tanto en los procesos de acreditación como en la evaluación de sexenios, se ha comenzado a valorar la disponibilidad en acceso abierto de las publicaciones, la existencia de *datasets* asociados y el uso de identificadores persistentes. El área de Apoyo a la Investigación del Servicio de Bibliotecas de la Universidad de Salamanca ha desarrollado campañas específicas de apoyo a estas convocatorias, incluyendo guías, talleres y servicios de revisión documental. Estos servicios se centran en ayudar a los investigadores a presentar sus datos de investigación de manera que cumplan con los requisitos de ANECA y maximicen su reconocimiento en los procesos de evaluación.

En el ámbito internacional, programas como Horizonte Europa exigen la elaboración de un DMP desde el inicio del proyecto, el depósito de los datos en repositorios FAIR y la inclusión de estrategias de difusión y explotación de los resultados. Asimismo, las evaluaciones de unidades de excelencia y centros de investigación incorporan indicadores relacionados con la gestión de datos, la transparencia y la apertura del conocimiento. Por eso se está trabajando para alinear los servicios bibliotecarios de apoyo a la investigación con las exigencias de los programas internacionales, ofreciendo asesoramiento y formación sobre la elaboración de DMP, la aplicación de los principios FAIR y la difusión de datos de investigación.

Los servicios bibliotecarios de la USAL han respondido a estas exigencias con una oferta integral que incluye:

- Asesoramiento en la redacción de DMP y metadatos.
- Selección de repositorios adecuados.
- Apoyo en la citación de *datasets*.
- Apoyo en la búsqueda y narración de evidencias de uso, lectura e impacto de los *datasets*.
- Formación en políticas de acceso abierto y ciencia abierta.

Este enfoque proactivo no solo mejora la competitividad de los proyectos presentados, sino que también fortalece la cultura institucional de apertura, calidad y responsabilidad científica. La USAL está promoviendo una cultura de gestión de datos de investigación que valora la calidad, la integridad y la reutilización de los datos.

### 3.3. SERVICIOS DE APOYO A LA INVESTIGACIÓN EN LA USAL: EL PAPEL DE LA BIBLIOTECA

#### 3.3.1. *Formación en gestión de datos*

La formación en gestión de datos de investigación (RDM) se ha convertido en un pilar fundamental de los servicios bibliotecarios de apoyo a la investigación. La USAL, consciente de la creciente importancia de la GDI, ha desarrollado un programa formativo integral dirigido a investigadores, personal técnico y estudiantes de posgrado.

Este programa abarca diversos aspectos de la (RDM), desde la planificación y la organización de los datos hasta su preservación y compartición. Los contenidos se adaptan a las necesidades específicas de cada grupo de usuarios y se imparten en diferentes formatos, incluyendo talleres presenciales, cursos en línea y seminarios especializados. Se ha puesto un énfasis especial en la formación práctica, con ejercicios y casos de estudio que permiten a los participantes aplicar los conocimientos adquiridos a sus propios datos.

Entre los temas que se abordan en la formación se encuentran:

**Principios FAIR:** Se explican los principios FAIR (Findable, Accessible, Interoperable, Reusable) y su importancia para la gestión y el intercambio de datos. Se proporcionan ejemplos prácticos de cómo aplicar estos principios en diferentes contextos de investigación. Se discuten las implicaciones de los principios FAIR para la selección de repositorios, la creación de metadatos y la elección de licencias de datos.

**Planes de Gestión de Datos (DMP):** Se ofrece orientación sobre la elaboración de DMP, incluyendo la identificación de los datos a generar, las estrategias de almacenamiento y preservación, y los aspectos éticos y legales a considerar. Se proporcionan plantillas y ejemplos de DMP para diferentes disciplinas. Se aborda la importancia de los DMP para la financiación de la investigación y se ofrecen consejos para la elaboración de DMP que cumplan con los requisitos de los financiadores.

**Metadatos:** Se explica la importancia de los metadatos para la descripción y el descubrimiento de los datos. Se presentan los principales estándares de metadatos y se ofrecen herramientas para su creación y gestión. Se discuten las diferencias entre los diferentes estándares de metadatos y se ofrece orientación sobre la selección del estándar más adecuado para cada tipo de dato.

**Repositorios de datos:** Se informa sobre los diferentes tipos de repositorios de datos disponibles y se proporcionan criterios para la selección del repositorio más adecuado para cada tipo de dato. Se explica cómo depositar los datos en repositorios y cómo obtener un identificador persistente (DOI) para los mismos. Se discuten las ventajas y las desventajas de los diferentes tipos de repositorios y se ofrece asesoramiento sobre la elección del repositorio más adecuado para cada disciplina.

Licencias de datos: Se presentan las diferentes licencias de datos disponibles y se explican sus implicaciones para la reutilización de los datos. Se ofrece orientación sobre la selección de la licencia más adecuada para cada tipo de dato. Se discuten las implicaciones legales y éticas de las diferentes licencias de datos y se ofrece asesoramiento sobre la elección de la licencia que mejor proteja los derechos de los investigadores y facilite la reutilización de los datos.

La formación en RDM no se limita a la transmisión de conocimientos teóricos, sino que también incluye actividades prácticas y ejercicios que permiten a los participantes aplicar los conceptos aprendidos a sus propios datos. Además, se fomenta el intercambio de experiencias y buenas prácticas entre los participantes. Se crean espacios para la discusión y el debate sobre los desafíos y las oportunidades en la gestión de datos de investigación.

Desde el Servicio de Bibliotecas se evalúa periódicamente la eficacia de la formación en RDM y realiza ajustes en el programa en función de las necesidades y demandas de la comunidad investigadora. Se utilizan encuestas, entrevistas y grupos de discusión para recopilar información sobre la satisfacción de los participantes y el impacto de la formación en sus prácticas de gestión de datos.

### *3.3.2. Asesoramiento en Planes de Gestión de Datos (DMP)*

Los Planes de Gestión de Datos (DMP) se han convertido en un requisito fundamental para la financiación de proyectos de investigación en muchos programas nacionales e internacionales. Un DMP es un documento vivo que describe cómo se van a gestionar los datos generados o recopilados durante un proyecto de investigación.

El Servicio de Bibliotecas de la USAL ofrece un servicio de asesoramiento especializado en la elaboración de DMP. Este servicio tiene como objetivo apoyar a los investigadores en la planificación y la organización de sus datos, garantizando que cumplan con los requisitos de los financiadores y que se ajusten a las buenas prácticas en RDM.

El servicio de asesoramiento en DMP incluye:

- Orientación sobre los requisitos de los financiadores: Se informa a los investigadores sobre los requisitos específicos de cada programa de financiación en relación con los DMP. Se proporcionan guías y plantillas específicas para cada financiador. Se ofrece asesoramiento sobre la interpretación de los requisitos de los financiadores y se aclaran las dudas que puedan surgir.
- Apoyo en la elaboración del DMP: Se proporciona orientación sobre la estructura y el contenido del DMP, incluyendo la identificación de los datos a generar, las estrategias de almacenamiento y preservación y los aspectos éticos y legales a considerar. Se ofrecen ejemplos de DMP exitosos y se proporcionan consejos para la redacción de un DMP claro y conciso.

- Revisión y *feedback* sobre el DMP: Se revisan los DMP elaborados por los investigadores y se proporcionan comentarios y sugerencias de mejora. Se ofrece asesoramiento sobre la resolución de los problemas identificados en el DMP.
- Herramientas y plantillas: Se ponen a disposición de los investigadores herramientas y plantillas para facilitar la elaboración del DMP. Se ofrecen recursos en línea y se proporciona acceso a software para la gestión de datos.

El servicio de asesoramiento en DMP se adapta a las necesidades específicas de cada proyecto de investigación y se ofrece de forma personalizada. Los bibliotecarios que prestan este servicio cuentan con una formación especializada en RDM y están al día de las últimas novedades en este campo. Se realizan cursos de formación continua y se participa en conferencias y talleres para mantenerse actualizado sobre las mejores prácticas en la gestión de datos de investigación.

### 3.3.3. Repositorio institucional GREDOS

El repositorio institucional GREDOS de la Universidad de Salamanca desempeña un papel crucial en la preservación y la difusión de la producción científica de la institución, incluyendo los datos de investigación. GREDOS proporciona una plataforma para el almacenamiento seguro y a largo plazo de los datos, así como para su acceso y reutilización por parte de la comunidad investigadora y la sociedad en general.

GREDOS cumple con los estándares internacionales de interoperabilidad y utiliza metadatos que facilitan el descubrimiento y la recuperación de los datos. Además, asigna identificadores persistentes (DOI) a los *datasets*, lo que garantiza su citabilidad y facilita su seguimiento. Se han implementado medidas de seguridad para proteger los datos de la pérdida, el daño y el acceso no autorizado. El repositorio está siendo recolectado por R3data, el registro internacional de repositorios de datos de investigación, lo que asegura la interoperabilidad y la visibilidad de su contenido.

En la continua búsqueda para ofrecer los mejores servicios, se han materializado varios convenios con instituciones externas, es el caso del acuerdo para el uso de la tecnología de Supercomputación Castilla y León (SCAYLE) —entidad pública creada por la Junta de Castilla y León y la Universidad de León, que tiene por objeto la mejora de las tareas de investigación de la universidad, los centros de investigación y las empresas de Castilla y León— para poder almacenar los datos cuyo peso digital excede el soportado por el repositorio GREDOS.

El depósito de datos en GREDOS ofrece numerosas ventajas para los investigadores, entre las que se encuentran:

- Mayor visibilidad e impacto: Los datos depositados en GREDOS están disponibles para una audiencia global, lo que aumenta su visibilidad y su potencial de impacto (Ferrerías-Fernández *et al.*, 2016). Se utilizan herramientas de análisis de uso y citación para medir el impacto de los datos depositados en GREDOS.
- Preservación a largo plazo: GREDOS garantiza la preservación a largo plazo de los datos, asegurando que estén disponibles para futuras investigaciones. Se implementan

estrategias de preservación digital para garantizar la integridad y la accesibilidad de los datos a lo largo del tiempo.

- Cumplimiento de los requisitos de los financiadores: Muchos financiadores exigen el depósito de los datos en repositorios de acceso abierto. GREDOS permite cumplir con estos requisitos. Se ofrece asesoramiento sobre los requisitos de los diferentes financiadores en relación con el depósito de datos.
- Citabilidad de los datos: La asignación de DOI a los *datasets* permite su citación en publicaciones científicas, lo que contribuye al reconocimiento del trabajo de los investigadores. Se ofrecen guías y herramientas para la citación adecuada de los datos.

El Servicio de Bibliotecas de la USAL ofrece apoyo a los investigadores en el depósito de sus datos en GREDOS, incluyendo la preparación de los metadatos y la selección de la licencia de datos adecuada. Se proporcionan guías y tutoriales para facilitar el proceso de depósito y se ofrece asistencia personalizada a los investigadores que lo necesiten.

#### 3.3.4. Apoyo en metadatos

Los metadatos desempeñan un papel fundamental en la gestión de datos de investigación, ya que proporcionan información esencial sobre los datos que facilita su descubrimiento, comprensión, interoperabilidad y reutilización. El Servicio de Bibliotecas de la USAL ofrece apoyo a los investigadores en la creación y la gestión de metadatos, garantizando que estos cumplan con los estándares y las buenas prácticas en la materia.

El apoyo en metadatos incluye:

- Orientación sobre los estándares de metadatos: Se informa a los investigadores sobre los principales estándares de metadatos utilizados en diferentes disciplinas y se ofrece orientación sobre la selección del estándar más adecuado para cada tipo de dato. Se proporcionan ejemplos de metadatos para diferentes disciplinas y se ofrecen recursos en línea sobre los estándares de metadatos.
- Herramientas de creación de metadatos: Se ponen a disposición de los investigadores herramientas para la creación y la edición de metadatos. Se ofrecen editores de metadatos, generadores automáticos de metadatos y sistemas de gestión de metadatos.
- Revisión y validación de metadatos: Se revisan los metadatos creados por los investigadores y se validan para garantizar su calidad y conformidad con los estándares. Se ofrecen servicios de revisión de metadatos y se proporciona *feedback* a los investigadores.
- Formación en metadatos: Se imparten talleres y cursos sobre la creación y la gestión de metadatos. Se ofrecen cursos en línea y se proporcionan materiales de formación sobre los metadatos.

El Servicio de Bibliotecas de la USAL promueve el uso de metadatos ricos y descriptivos, que permitan a los usuarios comprender el contexto, la procedencia y las características de los datos. Además, se fomenta el uso de vocabularios controlados y ontologías para mejorar la

interoperabilidad de los datos. Se ofrecen herramientas y recursos para la creación de metadatos interoperables y se proporciona asesoramiento sobre la aplicación de vocabularios controlados y ontologías.

### 3.3.5. *Apoyo en métricas de investigación*

Las métricas de investigación desempeñan un papel cada vez más importante en la evaluación de la producción científica y el impacto de la investigación. El Servicio de Bibliotecas de la USAL ofrece apoyo a los investigadores en la comprensión y el uso de diferentes métricas, incluyendo las métricas bibliométricas tradicionales y las altmétricas.

El apoyo en métricas de investigación incluye:

- Información sobre las diferentes métricas: Se informa a los investigadores sobre las diferentes métricas disponibles, incluyendo sus fortalezas y limitaciones. Se proporcionan definiciones y explicaciones de las diferentes métricas y se ofrecen recursos en línea sobre las métricas de investigación.
- Herramientas de análisis de métricas: Se proporcionan herramientas para la recopilación y el análisis de métricas. Se ofrecen bases de datos bibliográficas, software de análisis de citas y plataformas de altmétricas.
- Asesoramiento en la interpretación de las métricas: Se ofrece asesoramiento sobre la interpretación de las métricas y su uso en la evaluación de la investigación. Se proporciona orientación sobre el uso responsable y crítico de las métricas y se ofrecen ejemplos de buenas prácticas en la interpretación de las métricas.
- Formación en métricas de investigación: Se imparten talleres y cursos en línea y se proporcionan materiales de formación sobre las métricas de investigación.

El Servicio de Bibliotecas de la USAL promueve el uso responsable y crítico de las métricas, evitando su uso como único criterio para la evaluación de la investigación. Además, se fomenta el uso de una variedad de métricas que permitan obtener una visión más completa del impacto de la investigación. Se ofrecen servicios de análisis de métricas personalizados y se proporciona asesoramiento sobre la selección de las métricas más adecuadas para cada contexto de investigación.

## 4. DESAFÍOS Y BUENAS PRÁCTICAS

La implementación de servicios de apoyo a la investigación en el ámbito de la gestión de datos, la formación y las métricas ha supuesto un importante desafío para el Servicio de Bibliotecas de la USAL. Se han identificado diversos obstáculos que han dificultado el desarrollo y la prestación de estos servicios.

Uno de los principales desafíos es la falta de concienciación y alfabetización en datos entre la comunidad investigadora. Muchos investigadores desconocen la importancia de la gestión de datos y no cuentan con las competencias necesarias para llevarla a cabo de forma eficaz. Esto se traduce en una baja demanda de los servicios ofrecidos por la biblioteca y en dificultades para su adopción. Para abordar este desafío, la USAL ha implementado campañas de sensibilización y programas de formación que buscan aumentar la concienciación sobre la importancia de la gestión de datos y proporcionar a los investigadores las habilidades necesarias para gestionar sus datos de forma eficaz. Se han utilizado diferentes canales de comunicación, como correos electrónicos, redes sociales y presentaciones en conferencias, para llegar a la comunidad investigadora.

Otro desafío importante es la heterogeneidad disciplinar en la gestión de datos. Las prácticas y los estándares de gestión de datos varían considerablemente entre las diferentes disciplinas, lo que dificulta el desarrollo de servicios genéricos que se adapten a las necesidades de todos los investigadores. Para superar este desafío, la USAL ha adoptado un enfoque flexible y adaptable, que permite ajustar los servicios a las necesidades específicas de cada disciplina. Esto implica la colaboración con expertos de diferentes áreas para desarrollar guías y recursos específicos para cada disciplina. Se han creado grupos de trabajo interdisciplinarios para compartir conocimientos y experiencias sobre la gestión de datos en diferentes áreas.

Además, el Servicio de Bibliotecas de la USAL se enfrenta a limitaciones estructurales y de personal especializado. La falta de recursos humanos y materiales dificulta la ampliación y la mejora de los servicios ofrecidos. Para mitigar estas limitaciones, la USAL ha buscado financiación externa y ha establecido alianzas con otras instituciones y organizaciones. También se ha invertido en la formación y el desarrollo profesional del personal bibliotecario, para mejorar sus competencias en la gestión de datos y otras áreas relevantes. Se han organizado cursos de formación específicos para el personal bibliotecario y se han establecido programas de mentoría para facilitar la transferencia de conocimientos y habilidades.

A pesar de estos desafíos, el Servicio de Bibliotecas de la USAL ha implementado una serie de buenas prácticas para superar los obstáculos y garantizar la eficacia de los servicios de apoyo a la investigación.

Entre estas buenas prácticas se encuentran:

- Colaboración con otras unidades de la universidad: El Servicio de Bibliotecas de la USAL colabora estrechamente con otras unidades de la universidad, como los servicios de informática, las unidades de gestión de la investigación y los centros de investigación, para ofrecer un apoyo integral a los investigadores. Esta colaboración permite aprovechar los recursos y la experiencia de diferentes áreas y garantizar que los servicios de apoyo a la investigación se integren de forma efectiva en el flujo de trabajo de los investigadores. Se han establecido comités y grupos de trabajo conjuntos para coordinar las actividades y evitar la duplicación de esfuerzos.
- Adaptación de los servicios a las necesidades de los usuarios: El Servicio de Bibliotecas de la USAL adapta sus servicios a las necesidades específicas de los investigadores, ofreciendo formación y asesoramiento personalizados. Esto implica la realización de encuestas y entrevistas para comprender las necesidades y los desafíos de los

investigadores y el desarrollo de servicios que respondan a estas necesidades. Se han implementado procedimientos de relaciones con los usuarios para gestionar las interacciones con los investigadores y garantizar un servicio de alta calidad.

- Promoción de la ciencia abierta: El Servicio de Bibliotecas de la USAL promueve los principios de la ciencia abierta, fomentando el acceso abierto a los datos y las publicaciones. Esto implica la participación en iniciativas de ciencia abierta a nivel nacional e internacional, y la promoción de prácticas de ciencia abierta entre los investigadores de la USAL. Se han organizado eventos y campañas de sensibilización para promover la ciencia abierta y se han desarrollado recursos educativos para informar a los investigadores sobre los beneficios y las implicaciones de la ciencia abierta, además de trabajar activamente en el documento de la política de ciencia abierta de la USAL.
- Evaluación continua de los servicios: El Servicio de Bibliotecas de la USAL evalúa continuamente la eficacia de sus servicios y realiza ajustes en función de los resultados de la evaluación. Esto implica la recopilación y el análisis de datos sobre el uso y el impacto de los servicios y la utilización de estos datos para mejorar la calidad y la relevancia de los servicios. Se han implementado sistemas de gestión de la calidad para garantizar la mejora continua de los servicios y se realizan auditorías periódicas para evaluar el cumplimiento de los estándares de calidad.

## 5. CONCLUSIONES

El estudio de caso de los servicios bibliotecarios de apoyo a la investigación de la USAL evidencia el papel fundamental que desempeñan las bibliotecas universitarias en el contexto de la ciencia abierta y la transformación digital.

El Servicio de Bibliotecas de la USAL ha desarrollado una oferta integral de servicios que abarca la formación en gestión de datos, el asesoramiento en planes de gestión, el apoyo en metadatos y métricas y la gestión del repositorio institucional. Estos servicios contribuyen a mejorar la calidad, la visibilidad y el impacto de la investigación, así como a garantizar el cumplimiento de los requisitos de los financiadores y las buenas prácticas en la gestión de datos.

Sin embargo, la implementación de estos servicios plantea importantes desafíos, como la falta de concienciación y alfabetización en datos, la heterogeneidad disciplinar y las limitaciones estructurales y de personal especializado.

Para superar estos desafíos, es necesario fortalecer la colaboración entre las bibliotecas y otras unidades de la universidad, adaptar los servicios a las necesidades de los usuarios, promover la ciencia abierta y realizar una evaluación continua de los servicios. Además, es fundamental invertir en la formación y el desarrollo profesional del personal bibliotecario, para garantizar que cuenten con las competencias necesarias para desempeñar su papel en el nuevo ecosistema de la investigación. Se deben establecer programas de formación continua y se deben proporcionar oportunidades para la participación en conferencias y talleres.

En definitiva, las bibliotecas universitarias, con equipos especializados y comprometidos, son actores clave para la calidad, la transparencia y el impacto de la investigación en el marco actual. Su papel va más allá de la mera provisión de información, y se extiende al apoyo activo a los investigadores en todas las etapas del proceso de investigación, desde la planificación y la gestión de los datos hasta la difusión y la evaluación de los resultados. Se deben fortalecer las alianzas entre las bibliotecas y otras unidades de la universidad para garantizar un apoyo integral a los investigadores.

## REFERENCIAS BIBLIOGRÁFICAS

- Agosti, M. y Zucchetto, P. (2018). The role of academic libraries in the management of research data: services and training for researchers. *Ibersid*, 12(1), 9-18.
- Biblioteca Nacional de España. (2020). *Estrategia y normas para la gestión de datos de investigación*.
- BOE. (2024). Resolución de 22 de mayo de 2024, de la Comisión Nacional Evaluadora de la Actividad Investigadora, por la que se publican los criterios específicos aprobados para la evaluación de la actividad investigadora. *Boletín Oficial del Estado*, 124, 705.
- Camón, M., Rey, B. y Balagué, N. (2012). Servicios de apoyo a la investigación en las bibliotecas universitarias catalanas: una aproximación. *El Profesional de la Información*, 21(6), 613-620.
- Cárdenas, M. A. (2016). La gestión de datos de investigación: un nuevo servicio de la biblioteca universitaria. *E-Ciencias de la Información*, 6(1).
- Corti, L., Van den Eynden, V., Bishop, L. y Woollard, M. (Eds.). (2014). *Managing and sharing research data: A guide to good practice*. Sage.
- Cox, A. M., Pinfield, S. y Smith, J. (2019). Moving towards openness? The experiences, attitudes and behaviours of academic librarians towards open access. *LIBER Quarterly*, 29(1).
- Dobrevá, M., Wilson, M., Woollard, M. y Brindley, L. (2010). *Developing research support services in academic libraries*. SCOUNL.
- Ferreras-Fernández, T., García-Peñalvo, F., Merlo-Vega, J. A. y Martín-Rodero, H. (2016). Providing open access to PhD theses: visibility and citation benefits. *Program*, 50(4). <https://doi.org/10.1108/PROG-04-2016-0039>
- González-Solar, L. (2018). Estudios de usuarios en bibliotecas de investigación: metodologías y técnicas. *Anales de Documentación*, 21(1), 1-18.
- Johnston, L. A. (Ed.). (2014). *Curating research data: From discovery to reuse*. ALA Neal-Schuman.
- Jones, S. (2011). *How to develop a data management and sharing plan*. DCC.
- REBIUN. (2023). *V Plan Estratégico REBIUN: 2024-2027*.
- Tenopir, C., Sandusky, R. J., Allard, S. y Birch, B. (2012). Research data management policy support. *Journal of Librarianship and Information Science*, 44(1), 5-14.
- Universidad de Salamanca. (2018). *Plan Estratégico de la Universidad de Salamanca 2018-2021*.
- Universidad de Salamanca. (2023). *Memoria de actividades 2022-2023*.
- Universidad de Salamanca. (2025). *Política de ciencia abierta*.
- Voutsás Márquez, J. (2012). *Guía para la gestión de datos de investigación*. UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información.



## INTEROPERABILIDAD Y ACCESO EN LOS REPOSITORIOS DE DATOS DE INVESTIGACIÓN DE LAS UNIVERSIDADES ESPAÑOLAS

### *INTEROPERABILITY AND ACCESS IN SPANISH UNIVERSITY RESEARCH DATA REPOSITORIES*

Ana B. RÍOS HILARIO<sup>1</sup> y Carlos DÍAZ-REDONDO<sup>2</sup>

<sup>1</sup>*Universidad de Salamanca. España*  
anarihi@usal.es

 <https://orcid.org/0000-0002-7052-5860>

<sup>2</sup>*Universidad de Salamanca. España*  
charlierdiaz@usal.es

 <https://orcid.org/0000-0002-1391-3429>

**RESUMEN:** Se realiza un análisis sobre el estado de interoperabilidad y acceso que presentan los repositorios de datos de investigación en las universidades españolas. Para ello, se diseñó un cuestionario que contiene los indicadores de la quinta sección de la *Guía para la evaluación de repositorios institucionales de investigación* elaborada en 2021 por la Fundación Española para la Ciencia y la Tecnología (FECYT). Este cuestionario se envió a 42 repositorios pertenecientes a universidades tanto públicas como privadas. Se obtuvieron finalmente 10 respuestas que cubrían, sin embargo, los repositorios de 17 instituciones. De sus respuestas se infiere el desempeño o no de los 17 criterios técnicos que han de cumplir los repositorios de acceso abierto para garantizar su solvencia y su interoperabilidad. A pesar del bajo índice de respuesta, los resultados obtenidos pueden ser hasta cierto punto extrapolables, ya que el 76 % de los repositorios universitarios emplean DSpace como programa de gestión y utilizan las mismas políticas de parametrización.

**PALABRAS CLAVE:** acceso abierto; interoperabilidad; repositorios de datos de investigación; universidades españolas.

**ABSTRACT:** An analysis of the interoperability and access status of research data repositories in Spanish universities is carried out. For this purpose, a questionnaire was designed containing the indicators of the fifth section of the *Guide for the evaluation of institutional research repositories* developed in 2021 by the Spanish Foundation for Science and Technology (FECYT). This questionnaire was sent to 42 repositories, belonging to both public and private universities. In the end, 10 responses

were obtained, covering, however, the repositories of 17 institutions. From their answers we can infer the performance or not of the 17 technical criteria that open access repositories must meet to ensure their solvency and interoperability. Despite the low response rate, the results obtained could be extrapolated to some extent, since 76 % of the university repositories use DSpace as their management program and apply the same parameterization policies.

KEYWORDS: open access; interoperability; research data repositories; Spanish universities.

## 1. INTRODUCCIÓN

El «acceso abierto» (Open Access) a la literatura científica se refiere a

su disponibilidad gratuita en la Internet pública, permitiendo a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o vincular los textos completos de estos artículos, rastrearlos para indexarlos, pasarlos como datos a un software o usarlos para cualquier otro propósito legal, sin barreras financieras, legales o técnicas más allá de las inseparables del acceso a Internet en sí. La única restricción a la reproducción y distribución, y la única función de los derechos de autor en este ámbito, debería ser dar a los autores el control sobre la integridad de su trabajo y el derecho a ser reconocidos y citados adecuadamente. (Declaración de Budapest, 2002)

Cuando se habla de ciencia abierta se plantea la difusión del conocimiento científico de manera libre, accesible en línea y reutilizable. Proponer un marco de ciencia abierta en las prácticas de la investigación se entiende bajo la promoción del acceso abierto a la investigación científica, a la reproductibilidad y a la evaluación en abierto. En este contexto aperturista, surge la necesidad y el interés por el acceso a los datos de investigación abiertos, que se encuentran vinculados a las mismas investigaciones científicas, puesto que los datos de investigación permiten la reproductibilidad y la evaluación del proceso, pero, también, permiten la reutilización científica (Monteagudo-Haro y Prieto-Gutiérrez, 2024).

La idea que subyace bajo el acceso abierto es que los resultados y los datos utilizados en investigaciones financiadas públicamente deberían estar disponibles para todos y tener licencias que permitan su reutilización por investigadores, empresas y ciudadanos (Barrueco *et al.*, 2021).

Aunque hay muchas definiciones de datos de investigación, el Grupo de Trabajo de Repositorios de Rebiun, en el documento *Metadatos básicos para la descripción de datos de investigación en los repositorios institucionales* (2018), los define «como todo aquel material generado o recogido durante el desarrollo de una investigación y que sirve como evidencia de esta. Los conjuntos de datos o *dataset* constituyen un recurso digital complejo, y, por ello, requieren una gestión y descripción en base a dos aspectos fundamentales, la normalización y la interoperabilidad».

De acuerdo con la OCDE, «datos de investigación» son los registros factuales (datos numéricos, textuales, imágenes y sonidos) utilizados como fuentes primarias para la investigación científica y que están aceptados por la comunidad investigadora como medios para validar

conclusiones. Un conjunto de datos de investigación o *dataset* constituye una representación sistemática y parcial del tema u objeto investigado.

Existen numerosas razones para compartir los datos de investigación. En el documento *Recomendaciones para la gestión de datos de investigación* (Melero, 2018), elaborado por la red Maredata, se detallan los siguientes motivos:

- Promover la innovación y la reutilización de los datos que potencialmente puedan tener nuevos usos.
- Facilitar la colaboración entre usuarios de datos, creadores de datos y reutilizadores.
- Maximizar la transparencia y la fiabilidad de los datos.
- Favorecer la reproducibilidad de los ensayos experimentales.
- Permitir la verificación de los resultados de investigación.
- Reducir costes al evitar la duplicación de datos.
- Aumentar el impacto y la visibilidad de la investigación.
- Promover los proyectos de investigación de los que provienen los datos y sus publicaciones.
- Generar un reconocimiento directo de los investigadores productores de datos, como ocurre con cualquier otro resultado de investigación.

Relacionado con la gestión de estos datos, se han definido los principios FAIR (Findable, Accessible, Interoperable, Re-usable) que hacen referencia a cómo deben tratarse los datos de investigación para que sean localizables, accesibles, interoperables y reutilizables (Australian Research Data Commons, 2025).

Figura 1. Principios FAIR



Fuente: Australian National Data Service. FAIR principles.

La herramienta online FAIR-Aware, desarrollada por el proyecto FAIRsFAIR (2021), ayuda a entender mejor los principios FAIR y cómo contribuyen a mejorar el valor y el impacto de los datos de investigación. Estos principios los desarrolla del siguiente modo:

*Findable.* Para poder reutilizar los datos de investigación primero hay que localizarlos. Algunos de los principios que se deben cumplir son:

- Se asignará un identificador único persistente a los datos.
- Se describirán los datos de forma completa usando metadatos.
- Los metadatos podrán ser recuperados o indexados en recursos de búsqueda (por ejemplo, motores de búsqueda) para aumentar su visibilidad.

*Accesible.* Una vez localizados los datos de investigación, es necesario saber cómo se va a poder acceder a los datos. Algunos de los principios que se deben cumplir son:

- Se recuperarán los datos mediante su identificador utilizando protocolos de comunicación estandarizados, abiertos e implementables. Estos protocolos permitirán realizar procesos de autenticación y autorización de acceso a los datos.
- Los metadatos estarán accesibles, aunque los datos no estén disponibles.

*Interoperable.* Es necesario que los datos puedan ser interoperables para que sean utilizados por otras aplicaciones o flujos de trabajo. Algunos de los principios que se deben cumplir son:

- Los metadatos usarán un lenguaje formal, accesible y ampliamente utilizado para la representación del conocimiento.
- Los metadatos mostrarán relaciones descriptivas detalladas entre diferentes recursos.

*Reusable.* La reutilización de los datos de investigación es el objetivo principal que se persigue con los principios FAIR. Algunos de los principios que se deben cumplir son:

- Los metadatos contendrán una licencia clara y accesible sobre la reutilización de los datos.
- El contenido de los datos (atributos, metodología, variables, software y hardware necesario, etc.) deberá ser descrito para facilitar la reutilización de los datos.

Teniendo en cuenta que la interoperabilidad es uno de los principios fundamentales que deben cumplir los datos de investigación, el objetivo principal de este trabajo es analizar la interoperabilidad y el acceso en los repositorios de datos de investigación de las universidades españolas. Partiendo de este propósito general, a través de este estudio se pretende alcanzar los siguientes objetivos de carácter específico:

- Identificar las universidades españolas, tanto públicas como privadas, que cuentan con un repositorio de datos de investigación.
- Analizar el grado de cumplimiento de los criterios estipulados en la sección 5, titulada «Interoperabilidad de los metadatos y acceso a los contenidos», de la *Guía para la evaluación de repositorios institucionales de investigación* (2021).
- Definir futuras líneas de investigación en relación con la interoperabilidad de los datos de investigación.
- Establecer una serie de propuestas de mejora en aras de un mejor cumplimiento de los criterios técnicos definidos.

Para cumplir con los objetivos propuestos, este texto se estructura en los siguientes apartados: en primer lugar, se realiza una introducción en la que se definen la ciencia abierta y los datos de investigación y donde se plantean los objetivos de la investigación. A continuación, se expone la metodología desplegada para alcanzarlos. Finalmente se detallan y discuten los resultados y, por último, se extraen una serie de conclusiones y propuestas de mejora.

## 2. METODOLOGÍA

Tras llevar a cabo la búsqueda e identificación de las fuentes de información necesarias para la realización de la introducción, se procedió a la selección de la muestra. Para ello, se acudió al sitio web de la Red de Bibliotecas Universitarias y Científicas Españolas (REBIUN), donde se localizaron 76 repositorios pertenecientes a universidades españolas. Entre ellos, 50 pertenecían a universidades de ámbito público, mientras que los 26 restantes correspondían al ámbito privado. De la totalidad de repositorios localizados, se realizó una criba tomando como criterio el alojamiento de datos de investigación. Así pues, en total se recopilaron 42 repositorios (Tabla 1). Cabe señalar que la gran mayoría, 39, pertenecían a instituciones públicas, mientras que las universidades privadas fueron tan solo 3 (Camilo José Cela, Mondragon Unibertsitatea y Universitat Oberta de Catalunya).

Tabla 1. Repositorios universitarios españoles con datos de investigación

#	Universidad	Repositorio
1	IE University	IE Zenodoo
2	Mondragon Unibertsitatea	eBiltegia
3	Universidad Autónoma de Madrid	Biblos-e Archivo
4	Universidad Camilo José Cela	Depósito Digital e-UCJC
5	Universidad Carlos III de Madrid	e-Archivo
6	Universidad Católica San Antonio de Murcia	RIUCAM
7	Universidad Complutense de Madrid	Docta Complutense
8	Universidad de Alcalá	e_Buah: Biblioteca Digital Universidad de Alcalá
9	Universidad de Alicante	RUA
10	Universidad de Almería	riUAL: Repositorio Institucional de la Universidad de Almería
11	Universidad de Burgos	RIUBU: Repositorio Institucional de la Universidad de Burgos
12	Universidad de Cantabria	UCrea: Repositorio abierto de la Universidad de Cantabria
13	Universidad de Córdoba	Helvia: Repositorio institucional de la Universidad de Córdoba
14	Universidad de Granada	Digibug: Repositorio Institucional de la Universidad de Granada

*Ana B. Ríos Hilario y Carlos Díaz-Redondo*  
*Interoperabilidad y acceso en los repositorios de datos de investigación de las universidades españolas*

#	Universidad	Repositorio
15	Universidad de Huelva	Arias Montano
16	Universidad de Málaga	RIUMA: Repositorio Institucional de la Universidad de Málaga
17	Universidad de Murcia	DIGITUM
18	Universidad de Salamanca	Gredos: Repositorio Institucional de la Universidad de Salamanca
19	Universidad de Sevilla	idUS: Depósito de investigación de la Universidad de Sevilla
20	Universidad de Valladolid	UVaDOC: Repositorio Documental de la Universidad de Valladolid
21	Universidad de Zaragoza	Zaguán
22	Universidad del País Vasco	ADDI
23	Universidad Internacional de La Rioja	Re-UNIR
24	Universidad Nacional de Educación a Distancia	e-Spacio
25	Universidad Pablo de Olavide	RIO: Repositorio Institucional Olavide
26	Universidad Politécnica de Cartagena	Repositorio Digital de la Universidad Politécnica de Cartagena
27	Universidad Politécnica de Madrid	Archivo Digital UPM
28	Universidad Rey Juan Carlos	BURJC-Digital
29	Universidad San Pablo CEU	CEU Repositorio Institucional
30	Universidade da Coruña	RUC: Repositorio Institucional da Universidade da Coruña
31	Universidade de Santiago de Compostela	Minerva: Repositorio institucional da USC
32	Universitat Autònoma de Barcelona	Dipòsit Digital de Documents de la UAB
33	Universitat de Barcelona	Dipòsit Digital de la Universitat de Barcelona
34	Universitat de Girona	DUGiDocs
35	Universitat de Lleida	Repositori Obert UdL
36	Universitat de València	RODERIC
37	Universitat Jaume I	Repositori UJI
38	Universitat Oberta de Catalunya	O2 Repositori UOC
39	Universitat Politècnica de Catalunya	UPCommons: Portal del coneixement obert de la UPC
40	Universitat Politècnica de València	RiuNet: Repositorio Institucional de la Universitat Politècnica de València
41	Universitat Pompeu Fabra	Repositori digital de la UPF
42	Universitat Rovira i Virgili	Repositori Institucional de la Universitat Rovira i Virgili

Fuente: Elaboración de los autores (2025).

Tras el establecimiento de la muestra objeto de estudio, se diseñó la metodología. En este caso, se ha tomado como referencia la *Guía para la evaluación de repositorios institucionales de investigación* elaborada por la Fundación Española para la Ciencia y la Tecnología (FECYT) en el año 2021. Esta herramienta tiene como finalidad principal «ser un instrumento de auditoría interna para mejorar la calidad de los repositorios, facilitar su proceso de indexación en RECOLECTA y su adaptación a los nuevos estándares establecidos por OpenAIRE (Open Access Infrastructure for Research in Europe)». El objetivo final es fortalecer los repositorios institucionales y asegurar su papel como infraestructuras tecnológicas esenciales para los procesos de implementación de las políticas de acceso abierto y ciencia abierta.

Esta guía se estructura en 8 secciones: «Visibilidad»; «Políticas»; «Aspectos legales»; «Metadatos descriptivos de la publicación»; «Interoperabilidad de los metadatos y acceso a los contenidos»; «Logs y estadísticas»; «Seguridad, autenticidad e integridad de los datos»; «Servicios y funcionalidades de valor añadido». Dado que no sería factible estudiar el conjunto de apartados en un estudio pensado para su divulgación en un foro científico, se ha optado por centrarnos en la sección denominada «Interoperabilidad de los metadatos y acceso a los contenidos», ya que consideramos que la interoperabilidad de todos estos repositorios y la unificación de criterios de calidad resultan ser de gran importancia para lograr el máximo aprovechamiento de la información que contienen.

Para la evaluación de este quinto apartado, la guía propone un total de 17 criterios técnicos que han de cumplir los repositorios de acceso abierto para garantizar su solvencia y su interoperabilidad. Estos indicadores evaluables son:

1. Se proveen los metadatos a través del protocolo OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting).
2. Identificación de recursos de investigación en el repositorio.
3. El servidor OAI-PMH dispone de esquemas de metadatos estándares.
4. En las cabeceras de las respuestas del servidor OAI-PMH se indican los espacios de nombres de los distintos esquemas de metadatos utilizados por el formato *oai\_oaire*.
5. Se marcan los registros eliminados.
6. El tiempo de vida del testigo de reanudación es de un mínimo de veinticuatro horas.
7. El correo electrónico del administrador del repositorio está disponible en la etiqueta *AdminEmail* dentro de la respuesta a una orden *Identify*.
8. Existe una declaración de *Description* en la respuesta a una orden *Identify*.
9. La entrega de registros a través del protocolo OAI-PMH es progresiva a través de lotes.
10. El tamaño de los lotes para la entrega de registros está dentro del rango de 100-500 registros.
11. El formato de la fecha expresado en la orden *Identify* coincide con el campo *datestamp* de los registros.
12. Integración con otros sistemas de información de la institución.
13. Inclusión de etiquetas *<meta...>* en las cabeceras HTML.
14. El repositorio implementa Schema.org para facilitar el acceso estructurado a sus metadatos.
15. El repositorio soporta otros protocolos y API para compartir metadatos y/o contenidos.

16. Uso extendido de identificadores persistentes (DOI, Handle, URN, ORCID, etc.).
17. El repositorio usa vocabularios controlados u ontologías cuyos conceptos están dotados de identificadores persistentes (PURL, DOI, URN...).

Tomando estos indicadores, se implementó un cuestionario a través de Google Forms. La valoración del cumplimiento de cada uno de los criterios definidos se estableció mediante respuestas cerradas con valores «Sí» o «No» que permitían valorar si el repositorio cumplía o no los indicadores. Este formulario se remitió a través de correo electrónico a los responsables de los repositorios, solicitando su colaboración para llevar a cabo la investigación. En concreto, se realizó un envío el 14/03/2025 y una llamada de recordatorio el 31/03/2025.

Obtuvimos 10 respuestas que cubrían los repositorios de 17 instituciones. En concreto, las instituciones participantes fueron: Universidad de Alicante, Universidad Complutense de Madrid, Universidad Católica de Murcia, Mondragon Unibertsitatea, Universitat Jaume I, Universidad de Huelva, Universidad Autónoma de Madrid, Universidad Carlos III de Madrid, Universidad de Córdoba y, finalmente, el CSUC (Consorti de Serveis Universitaris de Catalunya), que respondió el formulario en nombre de un grupo de universidades catalanas (Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat de Girona, Universitat Oberta de Catalunya, Universitat Pompeu Fabra, Universitat Rovira i Virgili, Universitat de Lleida y Universitat Politècnica de Catalunya).

Una vez obtenidos los resultados, se procedió a la tabulación de los datos, trabajando sobre la tabla de Excel generada por Google Forms.

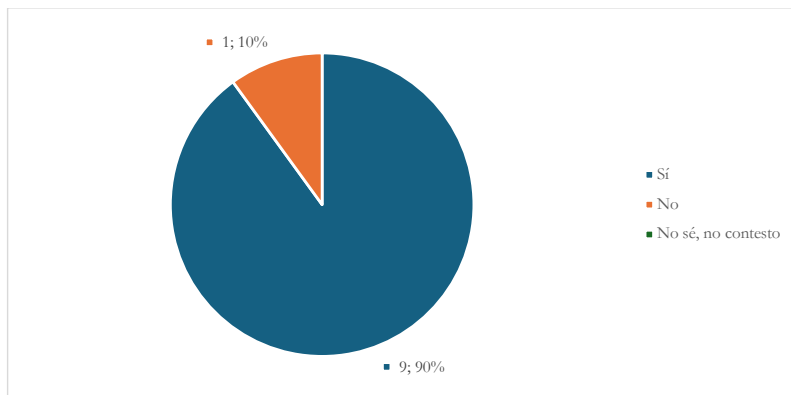
### 3. RESULTADOS Y DISCUSIÓN

Para comenzar, tenemos que decir que, de los 76 repositorios universitarios que conforman nuestra población objeto de estudio, 42 presentan un apartado destinado al depósito de los datos de investigación, frente a 34 que todavía no lo incluyen. Entre esos 42 repositorios con datos de investigación, 39 pertenecen a universidades públicas y 3 corresponden a universidades privadas. En este sentido, nos parece que este porcentaje debería ser más alto ya que una de las líneas estratégicas que promueve REBIUN es la de la biblioteca universitaria como «agente estratégico para el impulso de la ciencia abierta».

Con respecto al primer indicador evaluado, el 100 % de los encuestados afirman que los metadatos se proveen a través del protocolo OAI-PMH.

El segundo indicador hace referencia al modo de identificación de los recursos de investigación en los repositorios. Se ofrecen dos posibilidades: o bien que se realice través de un servidor OAI-PMH dedicado a investigación, o a través de uno o varios *sets* cuando el repositorio contenga materiales heterogéneos. En este caso, el 90 % de los encuestados responden que sí se cumple este requisito (Figura 2).

Figura 2. Modo de identificación de los recursos

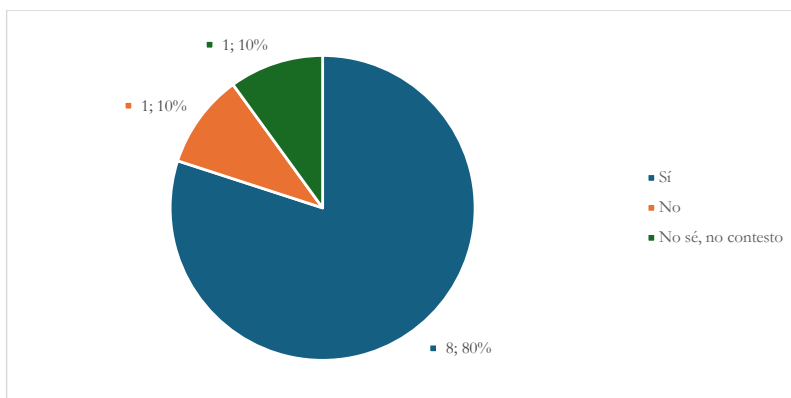


Fuente: Elaboración de los autores (2025).

El tercer indicador establece si el servidor OAI-PMH dispone de esquemas de metadatos estándares. Todos los repositorios emplean los esquemas mínimos de metadatos: oai\_dc y oai\_oaire. Este último seguirá el esquema descrito en las *OpenAire Guidelines for Literature Repository Managers 4.0*.

El cuarto indicador estudiado hace referencia a si en las cabeceras de las respuestas del servidor OAI-PMH se indican los espacios de nombres de los distintos esquemas de metadatos utilizados por el formato oai\_oaire. Aunque hay una mayoría de respuestas afirmativas (8), un encuestado opina de modo negativo y otro no contesta a esta pregunta (Figura 3).

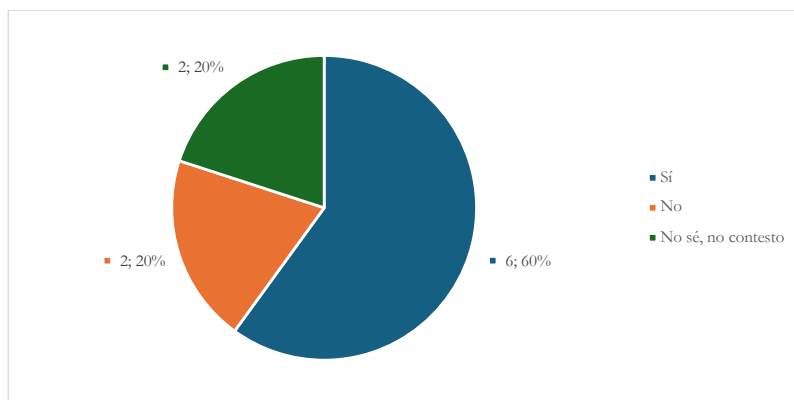
Figura 3. Indicación de los espacios de nombre en la cabecera



Fuente: Elaboración de los autores (2025).

El quinto indicador señala si se marcan los registros eliminados. El 60 % de los encuestados afirmaron que sí, mientras que el resto de los repositorios o bien no anotaban este dato o bien decidieron no contestar (Figura 4). En este sentido hay que comentar que «los registros eliminados deben marcarse durante, al menos, un periodo de tiempo suficiente que permita que los recolectores puedan identificarlos y eliminarlos de sus bases de datos. De otra forma se corre el riesgo de que registros eliminados por el proveedor de datos sigan existiendo en los recolectores» (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 22).

Figura 4. Identificación de los registros eliminados



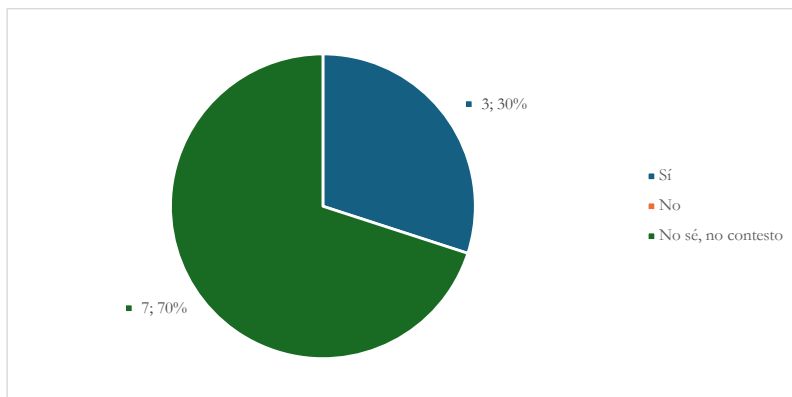
Fuente: Elaboración de los autores (2025).

En cuanto al sexto indicador, que refiere el tiempo de vida del testigo de reanudación, la guía establece un mínimo de veinticuatro horas. Es significativo que el 70 % (Figura 5) de los repositorios desconozca este dato, ya que, tal y como se apunta en los propios criterios, «los testigos de reanudación (*resumption token*) se utilizan en respuestas incompletas del servidor OAI-PMH para que el cliente pueda retomar la descarga en un momento posterior. La definición de hasta cuándo se puede retomar debe definirla cada repositorio, pero no podrá ser en ningún caso inferior a veinticuatro horas» (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 23).

Atendiendo al séptimo indicador, el 100 % de los encuestados contestan afirmativamente a la cuestión concerniente sobre si el correo electrónico del administrador del repositorio está disponible en la etiqueta AdminEmail dentro de la respuesta a una orden «*Identify*».

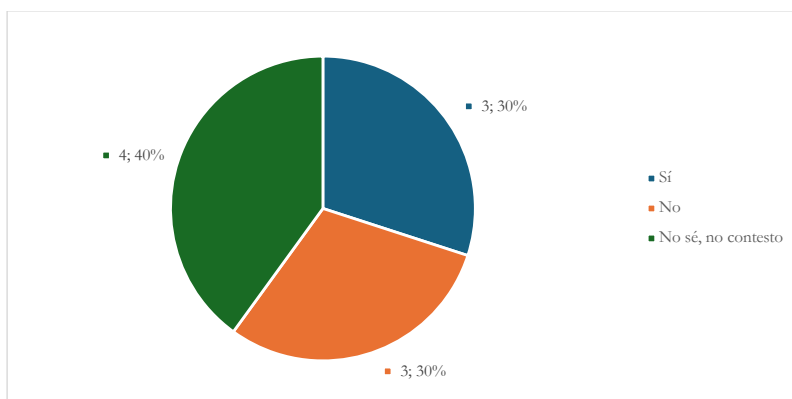
En el octavo indicador, que especifica si existe una declaración de «*Description*» en la respuesta a una orden «*Identify*», encontramos una mayor variedad de respuesta. Es así como tan solo un 30 % contesta afirmativamente. El resto del porcentaje se corresponde con el 40 % que no contesta y el 30 % que opina de modo negativo (Figura 6). Esta etiqueta se utiliza para describir el repositorio de una forma entendible por los recolectores. Existen diferentes esquemas ya definidos en el documento *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting*, si bien cada repositorio puede definir el suyo (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 23).

Figura 5. Tiempo de vida del testigo de reanudación



Fuente: Elaboración de los autores (2025).

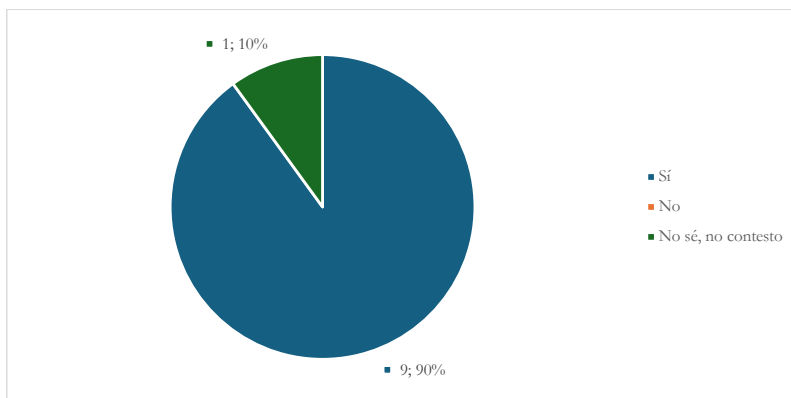
Figura 6. Declaración de «Description»



Fuente: Elaboración de los autores (2025).

El noveno indicador hace referencia a si la entrega de registros a través del protocolo OAI-PMH es progresiva a través de lotes. Tal y como se observa en el gráfico, de forma mayoritaria la entrega se produce de este modo (Figura 7).

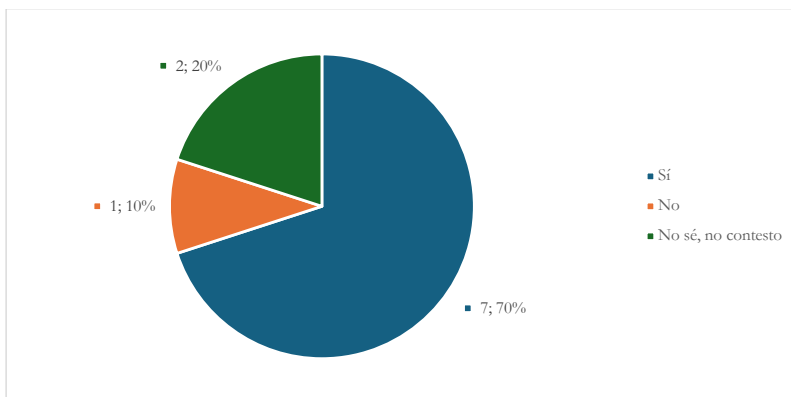
Figura 7. Entrega de registros progresiva a través de lotes



Fuente: Elaboración de los autores (2025).

Muy relacionado con este dato, señalamos el décimo indicador, que alude a la consulta relativa al tamaño de los lotes para la entrega de registros. Esta debe situarse dentro del rango de 100-500 registros. Está comprobado en la práctica que un número de elementos comprendido dentro de este rango agiliza los procesos de recolección y evita sobrecargas en los repositorios (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 23). El 70 % de los repositorios cumple con este criterio (Figura 8).

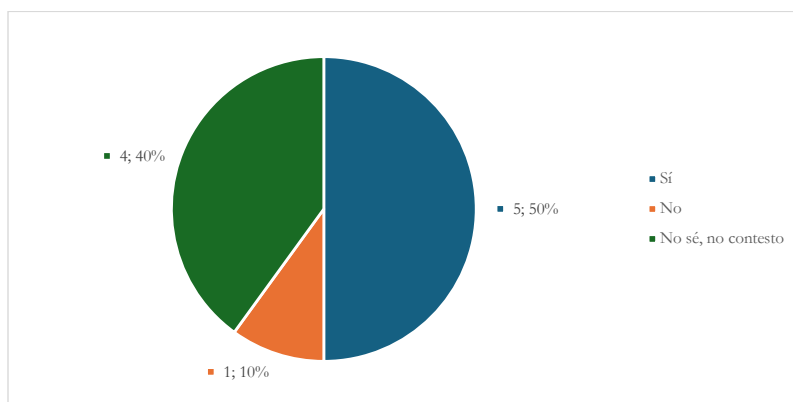
Figura 8. Tamaño de los lotes



Fuente: Elaboración de los autores (2025).

El undécimo indicador aborda si el formato de la fecha expresado en la orden «*Identify*» coincide con el campo «*datestamp*» de los registros (Figura 9). La mitad de los encuestados responden de modo positivo, mientras que un 40 % desconoce este dato y un caso contesta negativamente. Según las recomendaciones debe usarse el mismo formato para expresar la fecha de los registros que la definida en la granularidad de la página de identificación (*verb Identify*). La recolección gradual no es posible cuando la granularidad de entrada no puede ser procesada. Por lo tanto, el patrón definido en el campo «*granularity*» debe coincidir con el valor del campo «*datestamp*» de los registros. Es muy importante este hecho ya que este campo almacena la fecha y la hora específica en la cual ocurre un evento.

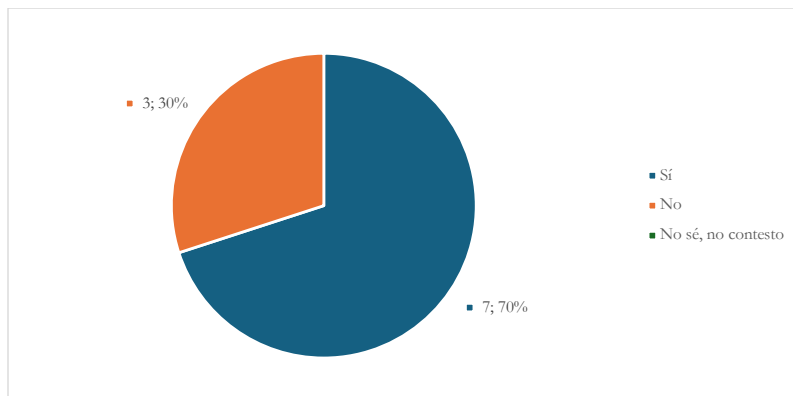
Figura 9. El formato de la fecha en la orden «*Identify*»



Fuente: Elaboración de los autores (2025).

El decimosegundo indicador indica la integración con otros sistemas de la información de la institución. En este caso destacamos que solamente 3 de los encuestados no cumplen con este requisito (Figura 10).

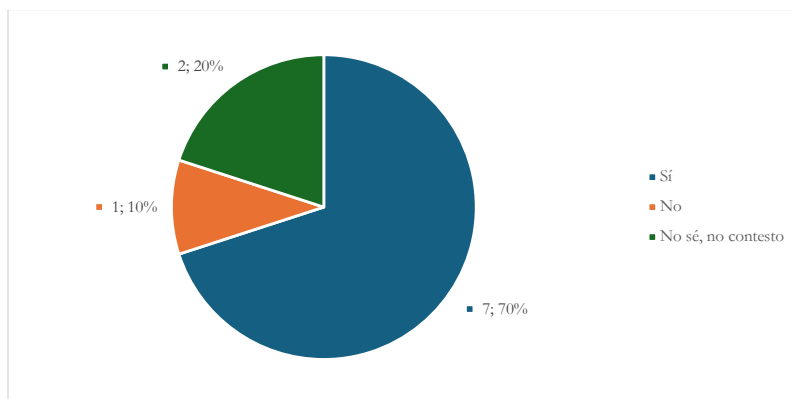
Figura 10. Integración con otros sistemas de la información



Fuente: Elaboración de los autores (2025).

El decimoterce indicador responde a la inclusión de las etiquetas <meta...> en la cabecera HTML. Estas etiquetas se utilizan para facilitar determinada información adicional a los motores de búsqueda y navegadores que atañe directamente al contenido de la página web. En 7 de los casos se da esta circunstancia, mientras que 2 afirman no saber o no responder (Figura 11).

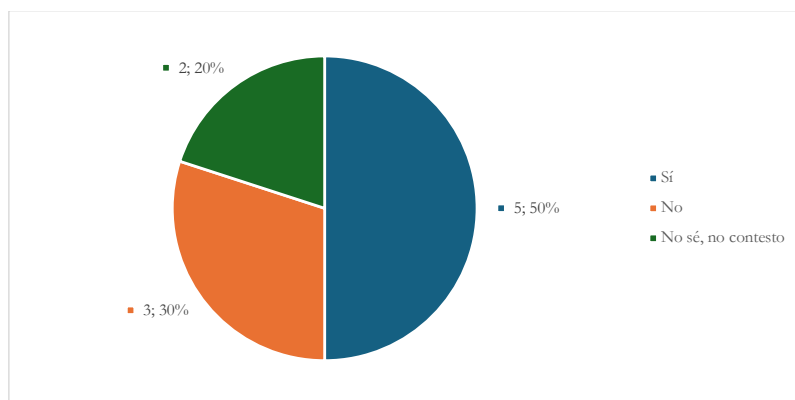
Figura 11. Inclusión de las etiquetas <meta...>



Fuente: Elaboración de los autores (2025).

Atendiendo al decimocuarto indicador, la implementación de Schema.org para facilitar el acceso estructurado a sus metadatos se refleja en la Figura 12. Schema.org es un vocabulario de datos estructurados que permite a los motores de búsqueda entender mejor el contenido de una página web, mejorando su representación en los resultados de búsqueda (SERP). Se contempla el etiquetado de los datos bibliográficos en las páginas web utilizando el modelo de marcado Schema.org (<https://schema.org/CreativeWork>) para facilitar su recuperación estructurada. Schema.org es ampliamente usado por motores de búsquedas comerciales como Google y cada vez más por agregadores de repositorios, por ejemplo, DataCite, por lo que su implementación facilita el descubrimiento y la accesibilidad a los contenidos de los repositorios (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 25). En este caso, hay que destacar que el 30 % de los repositorios encuestados no lo emplea, mientras que el 20 % directamente afirma desconocer este dato o no desea responder.

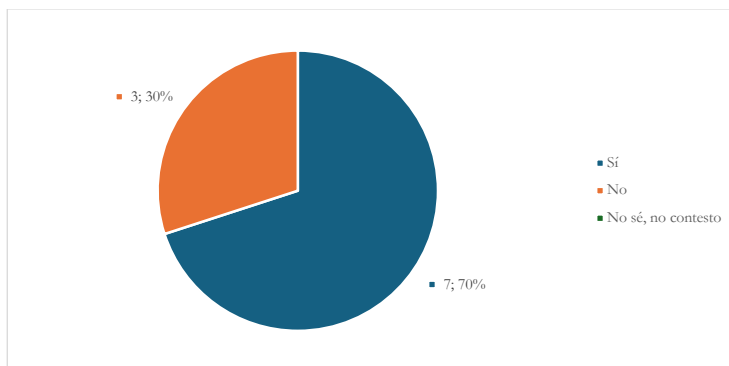
Figura 12. Implementación de Schema.org



Fuente: Elaboración de los autores (2025).

El indicador decimoquinto establece si se contemplan otros protocolos, aparte del básico OAI-PMH, para facilitar la recuperación de los metadatos y contenidos del repositorio (Figura 13). En los últimos años han surgido nuevos estándares que facilitan la interoperabilidad entre repositorios y otras infraestructuras afines. Estos estándares se encuadran dentro de los llamados servicios de la «nueva generación de repositorios» y permiten, entre otras funciones, la ingesta de contenidos (SWORD y API REST) y la sincronización de cambios asociados a metadatos y ficheros (ResourceSync) (*Guía para la evaluación de los repositorios institucionales*, 2021, p. 23). El 70 % de los repositorios de datos permiten otros protocolos.

Figura 13. Empleo de otros protocolos además de OAI-PMH

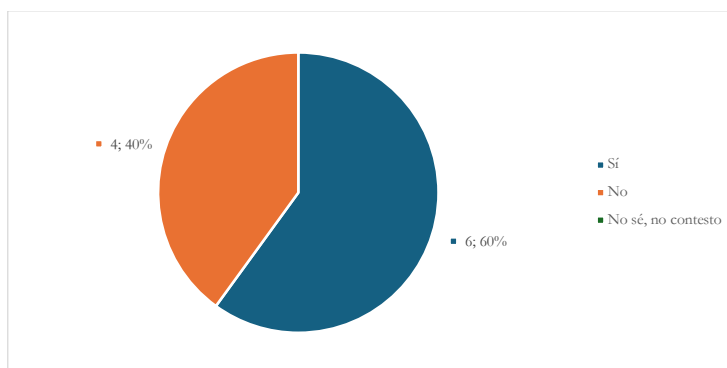


Fuente: Elaboración de los autores (2025).

El indicador decimosexto indaga sobre el uso extendido de identificadores persistentes tales como DOI, Handle, URN, ORCID, etc. Un 100 % de los repositorios consultados utilizan este tipo de identificadores. Tal y como consta en la *Guía para la evaluación de los repositorios institucionales* (2021, p. 25) es «una buena práctica incluir los identificadores persistentes de todas aquellas entidades, objetos y personas que se describen en los registros de metadatos de los repositorios ya que ello favorece una mayor interoperabilidad y facilita la explotación semántica de la información que albergan los contenidos de los repositorios».

Finalmente, en relación con este hecho, el indicador decimoséptimo inquiriría sobre si el repositorio emplea vocabularios controlados u ontologías cuyos conceptos están dotados de identificadores persistentes (PURL, DOI, URN...). En este caso las respuestas están más igualadas, ya que un 60 % los utiliza frente al 40 % que no lo hace (Figura 14).

Figura 14. Conceptos dotados de identificadores persistentes



Fuente: Elaboración de los autores (2025).

## 4. CONCLUSIONES

De los anteriores resultados podemos extraer las siguientes conclusiones. Aunque en los últimos años se ha dado un aumento en la constitución de repositorios de datos por parte de las universidades españolas se considera que, en este momento, todas las universidades deberían poseer un repositorio de datos de investigación. Recordemos que uno de los fines de las bibliotecas universitarias es su función como impulsoras de la ciencia abierta.

Por otro lado, si recurrimos al estudio realizado por Monteagudo-Haro y Prieto-Gutiérrez (2021), la presencia de un repositorio no involucra que este contenga datos de investigación. Del texto anterior se deduce que tan solo 6 repositorios presentaban más de 100 conjuntos de datos y casi 20 tienen menos de 15 conjuntos de datos de investigación, de los cuales 10 apenas alcanzaban 10 *datasets*.

Sobre el índice de respuesta debemos realizar varias consideraciones. En primer lugar, aunque el número de respuestas es bajo, tenemos que decir que el CSUC (Consorti de Serveis Universitaris de Catalunya) lo componen 8 repositorios, por lo que la muestra sería de 17 si cada uno de estos se hubiera contabilizado de modo individual. Sin embargo, los resultados obtenidos pueden ser extrapolables, ya que el 76 % de los repositorios universitarios utilizan DSpace como programa de gestión de repositorios. En este sentido, en líneas generales, utilizan los mismos parámetros y políticas de parametrización. Si a este dato sumamos que la mayoría de las respuestas son bastante homogéneas podemos decir que los resultados alcanzados son bastantes representativos.

No obstante, desde aquí realizamos un llamamiento a las diferentes instituciones públicas para que respondan a los cuestionarios en los que se solicita información con el fin de emprender pesquisas por parte de los investigadores no solo universitarios, sino de otros organismos. En ocasiones, como ha sucedido en este trabajo, el cuestionario es el único instrumento que permite obtener los datos para poder llevar a cabo las investigaciones.

El cumplimiento de las recomendaciones expuestas en la *Guía para la evaluación de repositorios institucionales de investigación* tiene como consecuencia primordial una mayor visibilidad y accesibilidad del contenido. Se deberían efectuar otros estudios que analizaran el resto de los criterios especificados en dicho instrumento.

Finalmente, se considera oportuno incluir un documento en el propio repositorio de datos que detalle el número de criterios RECOLECTA que cumple. Esta reseña proporcionaría una mayor credibilidad y fiabilidad al repositorio de investigación.

## REFERENCIAS BIBLIOGRÁFICAS

- Australian Research Data Commons. (2025). *What being FAIR means*. ARDC. <https://ardc.edu.au/resource-hub/making-data-fair/>
- Barruco, J. M., Rico-Castro, P. y Bonora Eve, L. V. (2021). *Guía para la evaluación de Repositorios institucionales de Investigación*. Conferencia de Rectores y Rectoras de las Universidades Españolas. <http://hdl.handle.net/20.500.11967/809>

- Budapest Open Access Initiative. (2002). *Declaración de Budapest*. BOAI. <https://www.budapestopenaccessinitiative.org/read/spanish-translation/>
- Fostering FAIR Data Practices in Europe (2021). *FAIR-Aware*. Amsterdam: <https://fairaware.dans.knaw.nl/>
- Lagoze, C. *et al.* (Eds.). (2005). *Implementation guidelines for the Open Archives Initiative Protocol for Metadata Harvesting*. Open Archives Initiative. <https://www.openarchives.org/OAI/2.0/guidelines.htm>
- Melero, R. (2018). *Recomendaciones para la gestión de datos de investigación*. Maredata. <https://digital.csic.es/bitstream/10261/173801/1/Maredata-recomendaciones-ESP.pdf>
- Monteagudo-Haro, P. y Prieto-Gutiérrez, J. J. (2024). Datos abiertos de investigación en repositorios universitarios españoles. *Revista Española de Documentación Científica*, 47(3), e397. <https://doi.org/10.3989/redc.2024.3.1581>
- OpenAIRE. (2018). *Guidelines for Literature Repository Managers*, vol. 4. OpenAIRE. <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/>
- Organisation for Economic Co-Operation and Development OECD (2007). *Principles and Guidelines for Access to Research Data from Public Funding*. OECD. [https://www.oecd.org/en/publications/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding\\_9789264034020-en-fr.html](https://www.oecd.org/en/publications/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr.html)
- Red de Bibliotecas Universitarias y Científicas Españolas. (2022). *Directorio de repositorios institucionales REBIUN*. REBIUN. <https://www.rebiun.org/>
- Red de Bibliotecas Universitarias. Grupo de Trabajo de Repositorio. *Metadatos básicos para la descripción de datos de investigación en los repositorios institucionales* (2018). Red de Bibliotecas Universitarias. Grupo de Trabajo de Repositorios. <http://hdl.handle.net/20.500.11967/639>


## AUTENTICIDAD DE LOS DATOS DE INVESTIGACIÓN: ANÁLISIS DE MODELOS DE METADATOS EN LOS REPOSITORIOS ESPAÑOLES DE RE3DATA

### *AUTHENTICITY OF RESEARCH DATA: ANALYSIS OF METADATA STANDARDS IN SPANISH REPOSITORIES AT RE3DATA*

Marieta MARKS LÖW<sup>1</sup> y Josefa GALLEGO LORENZO<sup>2</sup>

<sup>1</sup> *Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina. Brasil*

[marietalow@gmail.com](mailto:marietalow@gmail.com)

 <https://orcid.org/0000-0002-4442-9365>

<sup>2</sup> *Departamento de Patrimonio Artístico y Documental. Área de Biblioteconomía y Documentación. Universidad de León. España*

[josefa.gallego@unileon.es](mailto:josefa.gallego@unileon.es)

 <https://orcid.org/0000-0002-5218-517X>

**RESUMEN:** Este estudio analiza los modelos de metadatos utilizados en repositorios españoles de la plataforma RE3DATA y su principal objetivo es evaluar los mecanismos de verificación que garantizan la autenticidad de los datos de investigación en estos repositorios, analizando la implementación de estándares de metadatos. La autenticidad de los datos es un aspecto esencial para la gestión de datos de investigación en el contexto de la ciencia abierta y para la confiabilidad y la reproducibilidad de los resultados científicos, ya que los modelos de metadatos desempeñan un papel fundamental en la preservación digital. Para ello se ha aplicado un enfoque metodológico de carácter bibliográfico y exploratorio orientado al análisis comparativo de los modelos más utilizados. Se identificaron 63 repositorios españoles en el agregador de repositorios RE3DATA y los cinco modelos más utilizados fueron: Dublin Core, DataCite, OAI-ORE, DDI e ISO 19115. Los resultados demuestran que Dublin Core es el modelo más utilizado, a pesar de ser el que presenta más debilidades estructurales para garantizar la autenticidad. En el análisis se aprecia que hay diferencias significativas entre los cinco modelos y se sugiere crear directrices tanto nacionales como internacionales claras sobre la autenticidad en datos para garantizar la confiabilidad y la reutilización de los datos científicos a largo plazo. Nuestra principal aportación consiste en proporcionar una evaluación sobre las fortalezas y las debilidades de los modelos de metadatos en cuanto a la autenticidad,

con información relevante para la toma de decisiones de gestores y responsables de políticas científicas para la preservación digital y la gestión de datos de investigación.

PALABRAS CLAVE: autenticidad; datos de investigación; modelos de metadatos; principios FAIR; repositorios de datos.

ABSTRACT: This study analyzes the metadata models used in Spanish repositories on the RE3DATA platform. Its main objective is to evaluate the verification mechanisms that guarantee the authenticity of research data in these repositories, analyzing the implementation of metadata standards. Data authenticity is an essential aspect of research data management in the context of open science and for the reliability and reproducibility of scientific results, as metadata models play a fundamental role in digital preservation. To this end, a bibliographic and exploratory methodological approach was applied, aimed at the comparative analysis of the most widely used models. Sixty-three Spanish repositories were identified in the RE3DATA repository aggregator, and the five most widely used models were: Dublin Core, DataCite, OAI-ORE, DDI, and ISO 19115. The results show that Dublin Core is the most widely used model, despite being the one with the most structural weaknesses in terms of ensuring authenticity. The analysis shows that there are significant differences between the five models and suggests that clear national and international guidelines on data authenticity should be created to ensure the long-term reliability and reuse of scientific data. Our main contribution is to provide an assessment of the strengths and weaknesses of metadata models in terms of authenticity, with relevant information for decision-making by managers and scientific policy makers for digital preservation and research data management.

KEYWORDS: authenticity; data repositories; FAIR principles; metadata models; research data.

## 1. INTRODUCCIÓN

Vivimos en una sociedad «datificada» donde los datos digitalizados forman parte de nuestras vidas y cualquier operación en internet deja rastro de datos e información que, una vez analizados, pueden convertirse en conocimiento (Martín González e Iglesias Rodríguez, 2023). La ciencia siempre ha usado datos como materia prima para la investigación y el avance del conocimiento y estos datos pueden ser de diferente naturaleza y procedencia, como datos de investigación (ciencia abierta), datos públicos (*open data*) o datos gubernamentales. Autores como Abadal y Anglada (2020) sostienen que la ciencia abierta es un cambio de modelo en la manera de hacer ciencia que supone realizar con una visión «abierta» (*open*) todas las etapas o fases de la investigación científica (diseño, recolección de datos, revisión, publicación, etc.).

Describir y documentar los datos constituye una de las acciones más relevantes en el proceso de gestión, ya que permitirá conocer el origen de los datos, contextualizarlos para posteriormente preservarlos, visualizarlos y recuperarlos adecuadamente (Cunha *et al.*, 2024). Por ello, la autenticidad de los datos es esencial para garantizar la validez e implica que los datos sean originales, obtenidos de manera ética y sin alteraciones que comprometan su integridad.

Su autenticidad es esencial para garantizar la reproducción y la reutilización confiable de los datos.

La correcta gestión, reutilización y preservación de datos de investigación representa desafíos del ecosistema científico actual. El exceso de datos puede generar duda acerca de la confiabilidad de los datos, especialmente en un momento en que las redes sociales permiten que los datos de investigación sean utilizados para generar desinformación y contenido sensacionalista (Jurado, 2023). Por lo tanto, contar con repositorios de datos de investigación que utilicen modelos de metadatos capaces de garantizar la autenticidad de los datos es una necesidad cada vez más reconocida por la comunidad científica.

Para que los datos sean seguros deben estar en repositorios confiables que cumplan con los requerimientos FAIR y los principios TRUST de transparencia, responsabilidad, foco en el usuario, sostenibilidad y tecnología, componentes esenciales para definir repositorios de datos confiables (De Giusti, 2021).

En este contexto, los principios FAIR (Findable, Accessible, Interoperable, Reusable) y TRUST (Transparency, Responsibility, User Focus, Sustainability, Technology) se han convertido en marcos normativos clave para la evaluación y el diseño de repositorios confiables. FAIR establece los criterios mínimos para asegurar la disponibilidad y la reutilización de los datos, mientras que TRUST se centra en los atributos organizativos e infraestructurales que garantizan la confianza en los repositorios digitales. La autenticidad, aunque no explícita en ambos marcos, es transversal a sus componentes, particularmente en lo que respecta a la integridad y la trazabilidad de los datos.

La garantía de autenticidad es crucial para asegurar la reproducibilidad de los resultados de investigación, un principio fundamental de la ciencia abierta (Silva y Silva, 2020). Los datos de investigación, como bienes estratégicos del patrimonio científico, exigen modelos de metadatos que aseguren su autenticidad durante su ciclo de vida, particularmente en repositorios abiertos (Gallego Lorenzo *et al.*, 2024).

Este estudio tiene por objetivo evaluar los mecanismos de verificación de autenticidad en repositorios españoles de la plataforma RE3DATA, analizando la implementación de estándares de metadatos para preservación digital. En la metodología se comparan cinco modelos predominantes (Dublin Core, DataCite, OAI-ORE, DDI e ISO 19115), utilizando como referencia las directrices de RECOLECTA sobre integridad y autenticidad. RECOLECTA, junto con iniciativas internacionales como las OpenAIRE Guidelines, de los datos FAIR (Wilkinson *et al.*, 2016) y las recomendaciones de COAR, promueven buenas prácticas para repositorios digitales en aspectos como descubrimiento, acceso, reutilización e integridad de los contenidos.

Los metadatos desempeñan un papel fundamental en la preservación digital y en la garantía de autenticidad, al proporcionar información estructurada que describe, contextualiza y localiza recursos digitales, facilitando su recuperación, uso y gestión a largo plazo. La autenticidad se define como la propiedad de un objeto digital de corresponder a su estado original y de no haber sufrido modificaciones no documentadas, siendo este un aspecto central para la confiabilidad de los datos de investigación. La autenticidad, en términos archivísticos y documentales, se refiere a la propiedad de un objeto digital de ser lo que pretende ser y de no

haber sido alterado de manera no autorizada. Para que esta autenticidad sea comprobable, es indispensable que los datos estén acompañados de metadatos capaces de documentar su procedencia, registrar transformaciones, establecer relaciones lógicas entre versiones y permitir la verificación técnica mediante mecanismos como *checksums* o firmas digitales. De tal modo que los sistemas de metadatos deben incluir elementos que registren la procedencia de los datos, documenten todas las transformaciones aplicadas y establezcan mecanismos de verificación para asegurar su integridad (Duranti y Rogers, 2022; Pacheco *et al.*, 2023).

En el contexto español, iniciativas como RECOLECTA (Recolector de Ciencia Abierta) han establecido directrices para la evaluación de repositorios institucionales (FECYT, 2021), incorporando criterios específicos sobre seguridad, autenticidad e integridad de los datos. Estas directrices están alineadas con estándares internacionales como las OpenAIRE Guidelines y las recomendaciones de COAR (Confederation of Open Access Repositories), que promueven buenas prácticas para repositorios digitales en aspectos como descubrimiento, acceso, reutilización e integridad de los contenidos (COAR, 2022).

El objetivo principal del trabajo es ofrecer una evaluación comparativa de estos modelos de metadatos desde la perspectiva de la autenticidad digital, considerando tanto sus fortalezas como sus limitaciones en el contexto específico de los repositorios españoles. Se busca así contribuir a la discusión sobre buenas prácticas en la gestión de datos de investigación y proporcionar insumos para la toma de decisiones por parte de gestores de repositorios, desarrolladores de políticas científicas y la comunidad académica en general.

## 2. LA AUTENTICIDAD DE LOS DATOS EN REPOSITARIOS DIGITALES

La preservación digital se ha consolidado como un campo estratégico dentro de la Ciencia de la Información, orientado a asegurar el acceso a largo plazo de los recursos digitales mediante procesos técnicos, organizativos y normativos que salvaguardan su integridad, legibilidad y autenticidad. En este contexto, los principios FAIR y TRUST han sido adoptados ampliamente como marcos de referencia para evaluar la calidad y la sostenibilidad de los repositorios de datos de investigación.

Los principios FAIR –Findable, Accessible, Interoperable y Reusable–, formulados por Wilkinson *et al.* (2016), establecen que los datos deben estar adecuadamente descritos mediante metadatos; contar con identificadores persistentes y sistemas de recuperación; ser accesibles bajo condiciones claras, interoperables mediante vocabularios estandarizados, y reutilizables con información sobre su procedencia, licencias y contexto. La autenticidad se vincula estrechamente con los principios de interoperabilidad y reutilización, ya que la validez de los datos reutilizados depende de su integridad y trazabilidad documentada.

Por su parte, los principios TRUST –Transparency, Responsibility, User focus, Sustainability y Technology–, propuestos por Lin *et al.* (2020), aportan una perspectiva institucional y organizativa. TRUST promueve la existencia de políticas claras sobre preservación, estructuras de gobernanza responsables, atención a las necesidades de las comunidades usuarias, sostenibilidad financiera y adopción de tecnologías robustas. Aunque no establece criterios

técnicos específicos, este marco refuerza la necesidad de repositorios que integren prácticas orientadas a mantener la autenticidad documental de los datos científicos en contextos complejos y cambiantes.

En conjunto, FAIR y TRUST articulan una visión holística de la preservación digital que supera las preocupaciones técnicas aisladas e incorpora dimensiones éticas, legales y de gobernanza. La autenticidad de los datos, como atributo transversal, requiere no solo mecanismos técnicos, sino también marcos institucionales que garanticen su implementación coherente y sistemática.

La noción de autenticidad en entornos digitales ha sido objeto de creciente atención por parte de la archivística contemporánea, particularmente, desde el proyecto InterPARES (International Research on Permanent Authentic Records in Electronic Systems), que define la autenticidad como la confianza en que un documento es lo que declara ser y no ha sido alterado sin autorización durante su ciclo de vida (Duranti, 2007).

En el contexto de los datos de investigación, la autenticidad se articula a través de tres dimensiones principales: integridad, procedencia y fijeza (Duranti *et al.*, 2022). La integridad se refiere a la totalidad y la consistencia del objeto digital; la procedencia, a su origen y contexto de producción; y la fijeza, a su estabilidad material desde el punto de vista informático. Estos elementos deben estar documentados y ser verificables por medio de evidencias registradas en metadatos estructurados.

Entre los mecanismos más frecuentes para garantizar la autenticidad se encuentran los *checksums* (funciones *hash* que permiten verificar la integridad del archivo); las firmas digitales (que certifican la autoría y el no repudio), y, en escenarios más recientes, la tecnología *blockchain*, que permite registrar transacciones y cambios en cadenas inmutables, aunque su aplicación a gran escala en repositorios aún está en fase experimental.

Es importante destacar que la autenticidad digital no puede garantizarse de manera retrospectiva; requiere ser diseñada desde la planificación del ciclo de vida de los datos y sustentada en políticas, procedimientos y estructuras tecnológicas integradas. Por tanto, la elección del modelo de metadatos y su implementación rigurosa son elementos cruciales para esta garantía. Como señalan Kanza y Knight (2022), la gestión efectiva de los datos de investigación debe planificarse desde las primeras fases del proyecto, ya que la ausencia de dicha planificación puede comprometer la integridad y la reutilización futura de los datos. Igualmente ocurre con los datos insertados en los repositorios.

## 2.1. MODELOS DE METADATOS Y LA AUTENTICIDAD

Los metadatos, entendidos como datos sobre los datos, cumplen funciones esenciales en la preservación digital: describen contenido, contexto y estructura de los objetos digitales; documentan su historia de gestión, y facilitan su recuperación, interoperabilidad y autenticación. En el ámbito específico de la preservación, los metadatos deben incluir elementos que permitan rastrear la procedencia, registrar transformaciones, aplicar mecanismos de verificación y establecer relaciones entre versiones.

Existen diversos modelos de metadatos orientados a diferentes dominios científicos y técnicos. Los más ampliamente utilizados en repositorios de datos de investigación incluyen Dublin Core, DataCite y OAI-ORE (Khan *et al.*, 2024). Cada uno de estos esquemas incorpora, en distintos grados, elementos relevantes para la documentación de la autenticidad.

La literatura especializada ha destacado la necesidad de emplear modelos de metadatos que vayan más allá de la descripción bibliográfica básica e incorporen estructuras capaces de reflejar el ciclo de vida de los datos, sus agentes responsables, eventos de transformación, licencias, versiones y validaciones (Silva y Silva, 2021). Elementos como *provenance*, *version*, *checksum*, *data curator* y *lineage* son fundamentales para este propósito.

En este sentido, diversas organizaciones han desarrollado recomendaciones que orientan la implementación de metadatos con foco en la autenticidad. Las directrices de RECOLECTA (FECYT, 2021), por ejemplo, establecen que los repositorios deben contar con políticas que aseguren la integridad, la trazabilidad y la seguridad de los datos, y recomiendan el uso de estándares interoperables que incluyan elementos de verificación. Asimismo, las OpenAIRE Guidelines for Data Archives promueven la adopción de prácticas de metadatos alineadas con FAIR y que faciliten la documentación de la procedencia, las condiciones de acceso y los derechos de uso (OpenAIRE, 2022).

En el ámbito internacional, COAR (Confederation of Open Access Repositories) ha desarrollado un Framework of Good Practices que articula principios de interoperabilidad, sostenibilidad y calidad en la gestión de repositorios, insistiendo en la necesidad de documentar eventos de preservación, auditorías, cambios y autorizaciones como evidencia de la autenticidad (COAR, 2022).

Por lo tanto, el análisis de los modelos de metadatos debe considerar no solo su adopción generalizada, sino también su capacidad estructural para registrar la procedencia y garantizar la trazabilidad de los datos a través del tiempo. Esta perspectiva resulta clave para evaluar en qué medida los repositorios españoles cumplen con los requisitos internacionales de autenticidad y qué mejoras podrían ser implementadas para fortalecer su confiabilidad a largo plazo.

### 3. METODOLOGÍA

Este estudio adopta un enfoque metodológico de carácter exploratorio y bibliográfico, orientado al análisis comparativo de la capacidad de los modelos de metadatos implementados en repositorios españoles para garantizar la autenticidad de los datos de investigación. La naturaleza exploratoria se justifica por la escasez de investigaciones previas que aborden de manera sistemática esta dimensión en el contexto específico de España. Asimismo, la dimensión comparativa permite establecer contrastes entre distintos esquemas de metadatos, mientras que el enfoque bibliográfico garantiza la investigación en términos conceptuales y normativos a través de la literatura actualizada.

El proceso metodológico se desarrolló en las siguientes fases:

1. Identificación de repositorios españoles registrados en RE3DATA (Registry of Research Data Repositories), un registro global de repositorios de datos de investigación.

2. Recopilación de información sobre los modelos de metadatos implementados en cada repositorio identificado y selección de los modelos de metadatos más representativos según su prevalencia en los repositorios analizados, excluyendo aquellos que utilizan solamente un repositorio.
3. Evaluación de la conformidad de estos modelos con directrices nacionales e internacionales.

El análisis se centró en repositorios registrados en RE3DATA, por tratarse de un popular registro global de repositorios de datos de investigación, reconocido por su cobertura, fiabilidad y nivel de actualización. RE3DATA (Registry of Research Data Repositories) es una iniciativa internacional lanzada en 2012 con el objetivo de proporcionar un punto de acceso global y estandarizado de información sobre estos repositorios de datos de investigación. Surgió como un proyecto conjunto entre DataCite, la German Research Foundation (DFG) y varias instituciones académicas europeas, consolidándose posteriormente como un servicio de referencia en el ecosistema de la ciencia abierta (Pampel *et al.*, 2013). RE3DATA no solo funciona como registro, sino también como mecanismo de monitoreo de la evolución de los repositorios en términos de calidad, cobertura temática y cumplimiento de estándares.

Para el estudio se realizó una selección de todos los repositorios con sede institucional en España, identificando sus características, políticas de metadatos y cumplimiento con estándares internacionales.

A partir del total de 63 repositorios españoles registrados en RE3DATA (consulta realizada en febrero de 2024), se procedió a identificar los modelos de metadatos declarados por cada uno. Se contabilizaron un total de 10 modelos diferentes; para el análisis comparativo se seleccionaron solo aquellos que fueron utilizados por dos o más repositorios con el fin de garantizar una muestra representativa y evitar casos atípicos.

Los cinco modelos seleccionados fueron:

- Dublin Core (33 repositorios)
- DataCite Metadata Schema (7 repositorios)
- OAI-ORE (Open Archives Initiative - Object Reuse and Exchange) (4 repositorios)
- DDI (Data Documentation Initiative) (3 repositorios)
- ISO 19115 (2 repositorios)

Además del análisis documental de los modelos de metadatos y de los repositorios, se consultaron normativas relevantes como las OpenAIRE Guidelines for Data Archives, el Community Framework for Good Practices de COAR (2022) y la *Guía para la evaluación de repositorios institucionales* de RECOLECTA (FECYT, 2023), con el fin de establecer un marco normativo para la evaluación.

Cada modelo fue analizado con base en su estructura formal, sus elementos obligatorios y opcionales y su implementación concreta en los repositorios españoles identificados. Además, se consultaron perfiles de aplicación locales o extensiones institucionales para evaluar la personalización del modelo base.

Se reconocen varias limitaciones metodológicas. En primer lugar, la información declarada por los repositorios en RE3DATA puede no reflejar de forma completa o actualizada las prácticas reales de implementación de los esquemas de metadatos. En segundo lugar, no todos los repositorios disponen de documentación pública detallada, lo que obligó a inferir prácticas a partir de descripciones generales. Y, en tercer lugar, la evaluación no incluye análisis empírico de conjuntos de datos ni inspección directa de registros XML, lo que limita el alcance a una evaluación de tipo estructural y normativo.

Asimismo, el estudio no contempla aspectos relacionados con la gobernanza institucional, los recursos técnicos disponibles ni el grado de capacitación de los agentes responsables, factores que pueden incidir en la calidad de la implementación del modelo. Finalmente, la rápida evolución del ecosistema de ciencia abierta puede generar cambios significativos a corto plazo, por lo que se recomienda considerar estos resultados como una radiografía situada en un momento específico del desarrollo de los repositorios españoles.

## 4. RESULTADOS Y DISCUSIÓN

El análisis inicial de los datos levantados en el estudio reveló la existencia de 63 repositorios españoles registrados en RE3DATA, los cuales implementan un total de 10 modelos diferentes de metadatos. Aplicando el criterio de exclusión para aquellos modelos utilizados por un único repositorio, se seleccionaron cinco esquemas para un análisis en profundidad: Dublin Core (33 repositorios), DataCite Metadata Schema (7 repositorios), OAI-ORE - Open Archives Initiative Object Reuse and Exchange (4 repositorios), DDI - Data Documentation Initiative (3 repositorios) e ISO 19115 (2 repositorios). Este análisis ha permitido identificar diferencias significativas en los modelos de metadatos utilizados y, en particular, en su capacidad para garantizar la autenticidad de los datos de investigación. A continuación, presentamos un análisis detallado de cada modelo y una comparación entre los modelos en términos de su capacidad para garantizar la autenticidad de los datos en los repositorios de investigación.

### 4.1. DUBLIN CORE: FLEXIBILIDAD VERSUS ESPECIFICIDAD

El Dublin Core Metadata Initiative (DCMI) surgió en 1995 como un esfuerzo colaborativo orientado para crear un conjunto básico de elementos descriptivos que facilitara la interoperabilidad entre distintos sistemas de información. Originalmente concebido para describir recursos web, Dublin Core ha evolucionado hasta convertirse en uno de los estándares de metadatos más utilizados a nivel internacional, gracias a su simplicidad, flexibilidad y amplia adopción (DCMI, 2025).

La estructura original está compuesta por 15 elementos básicos, como *title*, *creator*, *subject*, *description* y *identifier*, que permiten representar de forma generalizada la información sobre un recurso digital. Estos elementos pueden ser extendidos o refinados mediante cualificadores, lo que ha dado lugar a numerosas adaptaciones locales y perfiles de aplicación.

En el contexto de los repositorios españoles registrados en RE3DATA, Dublin Core es el modelo predominante, implementado en 33 de los 63 repositorios analizados. Esta preferencia se explica por su bajo umbral técnico de implementación, compatibilidad con los protocolos OAI-PMH y amplia disponibilidad en plataformas de gestión como DSpace o EPrints, comúnmente empleadas por instituciones académicas.

No obstante, esta generalización ha derivado en una adopción frecuentemente limitada a los elementos más básicos del estándar, sin cualificaciones ni extensiones orientadas a la preservación digital o autenticidad de los datos de investigación.

Una de las principales fortalezas de Dublin Core es su alta interoperabilidad semántica y sintáctica, derivada de su estandarización por la ISO (ISO 15836) y su compatibilidad con múltiples formatos de serialización como XML y RDF. Su uso generalizado favorece la agregación de metadatos en catálogos nacionales e internacionales y facilita la integración en ecosistemas de ciencia abierta.

Además, su estructura simple lo convierte en una opción accesible para repositorios con recursos técnicos limitados o en etapas iniciales de desarrollo. También permite ser utilizado como núcleo sobre el cual pueden integrarse otros esquemas más especializados mediante esquemas compuestos o mapeos.

Pese a sus ventajas en términos de adopción e interoperabilidad, Dublin Core presenta carencias significativas en relación con la autenticidad digital. En su forma estándar, no contempla elementos explícitos para:

- Documentar la procedencia detallada de los datos (ej. agentes de recolección, transformación o validación).
- Registrar versiones, modificaciones o transformaciones del objeto digital.
- Implementar mecanismos de verificación técnica, como *checksums* o firmas digitales.
- Representar relaciones complejas entre múltiples objetos digitales interdependientes.

Algunos repositorios han intentado subsanar estas limitaciones mediante la incorporación de elementos adicionales o el uso de perfiles de aplicación locales. No obstante, estas soluciones no siempre están normalizadas ni documentadas públicamente, lo que compromete la interoperabilidad y dificulta la evaluación sistemática de su eficacia.

Entre los perfiles más relevantes se encuentran el Dublin Core Qualified y el uso de metadatos compuestos, integrando esquemas como PREMIS o METS. Sin embargo, la adopción de estos perfiles en repositorios españoles es aún incipiente y heterogénea, y no existe un consenso sobre su implementación mínima recomendada para contextos de preservación.

En conclusión, Dublin Core representa una solución práctica y ampliamente adoptada para la descripción básica de recursos, pero su capacidad para garantizar la autenticidad de los datos de investigación es limitada sin complementos técnicos y normativos adicionales. Esta situación plantea un desafío estratégico para los repositorios españoles que aspiran a alinearse con los estándares de calidad y fiabilidad exigidos por la ciencia abierta contemporánea.

#### 4.2. DATACITE METADATA SCHEMA: ROBUSTEZ EN LA VERIFICACIÓN

El DataCite Metadata Schema fue desarrollado por la organización DataCite como parte de su misión de facilitar la atribución, el descubrimiento y la citación de datos de investigación. Su objetivo principal es apoyar la asignación de identificadores persistentes (DOI) a conjuntos de datos y garantizar que estos estén acompañados por metadatos ricos, estructurados y alineados con principios FAIR (DataCite, 2021). A diferencia de Dublin Core, DataCite fue concebido específicamente para describir objetos digitales en el ámbito de la investigación científica, incorporando elementos diseñados para representar información clave sobre procedencia, integridad y versiones de los datos.

El esquema de DataCite incluye varios elementos particularmente relevantes para la verificación de la autenticidad de los datos, entre ellos:

- *version*: permite indicar la versión específica del conjunto de datos, facilitando la trazabilidad de modificaciones.
- *relatedIdentifiers*: establece relaciones entre objetos digitales, como versiones anteriores o archivos complementarios.
- *contributors* con roles específicos (ej. *DataCollector*, *DataCurator*, *DataManager*): documentan la cadena de custodia de los datos, permitiendo identificar responsables en distintas etapas del ciclo de vida.
- *formats* y *sizes*: ayudan a validar las características físicas del archivo.
- *description* con subtipos como *technicalInfo* o *methods*: ofrecen espacio para incluir documentación detallada de procesos técnicos o metodológicos.

Además, aunque no obligatorio, el esquema es compatible con la inclusión de *checksum* y *checksumType* como extensiones, lo que permite integrar mecanismos de verificación de integridad en la documentación del objeto digital.

Una de las contribuciones más significativas del esquema DataCite es su integración nativa con el sistema de identificadores persistentes DOI, lo cual asegura que cada conjunto de datos pueda ser referenciado de manera única, estable y verificable. Esta característica fortalece la trazabilidad y evita ambigüedades en la reutilización de datos. El DOI también actúa como ancla para metadatos actualizables, permitiendo reflejar cambios sin comprometer la integridad del identificador original.

A través de la estructura de *contributors* y del elemento *publisher*, el esquema permite registrar con claridad no solo quién creó los datos, sino también quién los recopiló, procesó, almacenó y curó. Esta capacidad para documentar la procedencia institucional y técnica de los datos es fundamental para verificar su autenticidad y evaluar su fiabilidad para usos secundarios.

En el contexto español, siete repositorios analizados en RE3DATA utilizan el esquema DataCite. Estos repositorios presentan una orientación más explícita hacia buenas prácticas de documentación y preservación. Algunos incluso integran DataCite con otros esquemas como PREMIS o METS, conformando estructuras de metadatos compuestas que fortalecen su capacidad para representar eventos, agentes y relaciones entre objetos digitales.

Aunque la implementación aún es limitada en número, los repositorios que adoptan DataCite muestran una tendencia a alinear sus políticas con los principios FAIR y con las directrices de iniciativas como RECOLECTA o OpenAIRE, incluyendo mecanismos de control de versiones, registro de transformaciones y documentación de integridad.

Comparado con Dublin Core, DataCite ofrece una estructura significativamente más rica y orientada a los desafíos actuales de la ciencia de datos. Su adopción favorece no solo la autenticidad, sino también la visibilidad y la interoperabilidad internacional. Se configura, así, como un modelo de referencia para repositorios que buscan elevar su nivel de calidad, trazabilidad y cumplimiento normativo.

#### 4.3. OAI-ORE: AGREGACIONES Y RELACIONES COMPLEJAS

El modelo OAI-ORE (Open Archives Initiative - Object Reuse and Exchange) fue desarrollado por la Open Archives Initiative para representar y describir agregaciones de recursos digitales distribuidos en la web. Su objetivo es permitir la especificación formal de conjuntos de objetos interrelacionados que, aunque almacenados por separado, forman una unidad lógica –como colecciones de datos, publicaciones científicas vinculadas a conjuntos de datos, o versiones sucesivas de un mismo archivo (Lagoze *et al.*, 2008)–.

OAI-ORE utiliza como base el modelo de datos Resource Description Framework (RDF) y serializaciones como RDF/XML o Atom. La unidad fundamental es el Resource Map (ReM), que describe la agregación (*Aggregation*) y sus recursos componentes, así como las relaciones entre ellos mediante propiedades RDF estándar o personalizadas.

El esquema OAI-ORE permite describir explícitamente relaciones semánticas y estructurales complejas, como versiones, derivados, documentación asociada, conjuntos de archivos o elementos multimedia vinculados. Además, ofrece la posibilidad de declarar metadatos sobre la agregación como un todo y sobre cada uno de sus componentes individualmente.

Esto convierte a OAI-ORE en una solución adecuada para contextos donde los datos de investigación no son archivos únicos, sino objetos compuestos distribuidos que requieren gestión colectiva y coherente.

De los 63 repositorios españoles identificados en RE3DATA, 4 implementan el modelo OAI-ORE. Estos repositorios suelen tener una orientación avanzada en cuanto a gestión de objetos digitales complejos, especialmente en disciplinas que generan conjuntos de datos estructurados, multimedia o vinculados a otras formas de documentación científica.

Sin embargo, el análisis revela que la implementación efectiva de OAI-ORE es aún incipiente y desigual. En algunos casos, el modelo se emplea como complemento de otros esquemas, sin aprovechar plenamente sus capacidades descriptivas. Además, no todos los repositorios proporcionan documentación pública sobre su uso específico de agregaciones, lo que dificulta evaluar su integración con prácticas de verificación de autenticidad.

Desde la perspectiva de la autenticidad digital, OAI-ORE presenta ventajas notables en cuanto a trazabilidad relacional. Su estructura permite declarar explícitamente qué recursos conforman un conjunto de datos, qué versiones o transformaciones han ocurrido y cómo se relacionan con otros objetos digitales. Esta capacidad de reconstruir el contexto documental

es esencial para establecer la autenticidad en entornos donde la procedencia y la organización interna del contenido son fundamentales.

#### 4.3.1. Desafíos de implementación

Pese a su potencial, OAI-ORE enfrenta varios desafíos para su adopción generalizada:

- Complejidad técnica: requiere conocimiento de RDF y tecnologías semánticas, lo que limita su implementación a instituciones con capacidades avanzadas.
- Falta de estandarización en perfiles de aplicación: la libertad estructural que ofrece el modelo puede generar inconsistencia semántica entre implementaciones.
- Escasa integración con mecanismos de verificación técnica como *checksums* o firmas digitales, los cuales deben ser implementados mediante esquemas adicionales.

OAI-ORE no fue diseñado originalmente como un modelo de preservación, pero su capacidad para describir relaciones y dependencias lo convierte en una herramienta valiosa para representar la procedencia y la estructura lógica de los datos. Integrado con otros modelos orientados a la verificación técnica, puede fortalecer significativamente la capacidad de los repositorios para garantizar la autenticidad de objetos digitales complejos.

#### 4.4. DDI: ESPECIALIZACIÓN EN CIENCIAS SOCIALES

La Data Documentation Initiative (DDI) es un modelo de metadatos desarrollado específicamente para describir datos utilizados en las ciencias sociales, en particular aquellos derivados de encuestas, estudios longitudinales, censos y otras formas de investigación empírica. Desde su primera versión en 2000, DDI ha evolucionado hacia una estructura modular y jerárquica capaz de describir todo el ciclo de vida de los datos, desde su recolección hasta su preservación y diseminación (Vardigan y Whiteman, 2007).

DDI ha sido promovido por un consorcio internacional de instituciones académicas y centros de datos, y cuenta con múltiples versiones (incluidas DDI-Codebook y DDI-Lifecycle), cada una adaptada a distintos niveles de complejidad y granularidad documental.

DDI destaca por su capacidad para representar la estructura interna, la historia de transformación y el contexto metodológico de los datos. El modelo permite describir:

- Instrumentos de recolección (ej. cuestionarios, entrevistas).
- Procesos de codificación, limpieza y edición.
- Versiones y modificaciones sucesivas.
- Información sobre anonimización y acceso.
- Procedencia institucional y técnica de cada etapa del ciclo.

A través de elementos como *Methodology*, *DataCollection*, *Processing*, *Version*, *Lineage* y *OtherMaterial*, DDI documenta con alta precisión cada evento asociado a la creación y la

modificación del conjunto de datos. Esta capacidad resulta crítica para establecer la autenticidad documental, especialmente en contextos donde la replicabilidad y la trazabilidad metodológica son esenciales.

Una de las características distintivas de DDI es su enfoque en el modelo de ciclo de vida completo. A diferencia de otros esquemas orientados principalmente a la descripción bibliográfica o técnica, DDI permite representar las transiciones temporales y estructurales que experimentan los datos desde su concepción hasta su archivo. Esta representación es esencial para validar que los cambios sufridos fueron autorizados, documentados y atribuibles a agentes identificados.

DDI ofrece mecanismos explícitos para asegurar la autenticidad, incluyendo:

- Registro detallado de transformaciones metodológicas (*processing steps*)
- Identificación de actores responsables por cada fase del proceso
- Versión del estudio y justificación de cambios
- Relación entre versiones y objetos derivados
- Información sobre restricciones de acceso y derechos de uso

Además, su diseño permite integrarse con otros esquemas como PREMIS o METS, incorporando funcionalidades adicionales para verificación técnica y preservación a largo plazo.

En el ámbito español, tres repositorios especializados en ciencias sociales utilizan DDI como modelo principal de metadatos. Estos repositorios están vinculados a instituciones académicas con tradición en investigación empírica y manejo de encuestas longitudinales. Su implementación de DDI suele estar alineada con estándares internacionales y con buenas prácticas documentales, incluyendo metadatos enriquecidos, versiones detalladas y políticas claras de curación de datos.

Aunque minoritaria en número, la presencia de DDI refleja un nivel de madurez documental elevado en estos repositorios, y sugiere un compromiso con la transparencia metodológica y la reproducibilidad científica.

DDI es uno de los pocos esquemas diseñados explícitamente para capturar la dimensión temporal, metodológica y contextual de los datos. Su potencial para garantizar la autenticidad es particularmente alto en disciplinas donde los datos son construidos mediante procesos complejos y donde la validación metodológica es parte del proceso científico.

Por tanto, se configura como un modelo de referencia para aquellos repositorios que gestionan datos cuantitativos estructurados, especialmente en áreas como sociología, economía, ciencia política y psicología.

#### 4.5. ISO 19115: ESTÁNDAR PARA INFORMACIÓN GEOGRÁFICA

La norma ISO 19115 forma parte de la familia de estándares ISO 19100, desarrollados por el comité técnico ISO/TC 211 para abordar las necesidades específicas de la información geoespacial. Publicada inicialmente en 2003 y actualizada en versiones posteriores (ISO 19115-1:2014; ISO 19115-2:2019), esta norma proporciona un marco integral para describir

recursos geográficos digitales, con especial énfasis en calidad, procedencia, referencia espacial y condiciones de uso (ISO, 2019).

Su adopción ha sido promovida por agencias nacionales de cartografía, instituciones académicas, infraestructuras de datos espaciales (SDI) y organismos internacionales como INSPIRE (en Europa) y GEOSS (a nivel global).

ISO 19115 se basa en un modelo de datos modular y extensible, articulado en torno a elementos centrales como identificación, calidad, distribución, referencia temporal y espacial y mantenimiento. Además, permite la descripción detallada de:

- Procedencia de los datos (*lineage*)
- Procesos de transformación y edición
- Restricciones legales y técnicas de uso
- Formatos y estructuras del archivo
- Contactos institucionales y técnicos responsables

El estándar se implementa comúnmente utilizando XML bajo el esquema Geographic Markup Language (GML) y se integra con otros estándares relacionados, como ISO 19139 (para serialización) e ISO 19157 (para evaluación de calidad).

#### 4.5.1. *Linaje y calidad de datos*

Uno de los componentes más relevantes de ISO 19115 para la autenticidad es el elemento LI\_Lineage, que documenta el historial completo de producción del recurso geoespacial. Incluye:

- Descripción de los procesos de obtención y transformación
- Identificación de fuentes utilizadas
- Agentes responsables de cada fase
- Métodos de verificación y control de calidad

Este enfoque estructurado proporciona evidencia detallada sobre la fiabilidad, la exactitud y la procedencia de los datos geográficos, facilitando su evaluación crítica y la reconstrucción de su historia documental.

En el análisis realizado, dos repositorios españoles especializados en información geográfica utilizan el modelo ISO 19115. Estos repositorios están asociados a centros de investigación o servicios cartográficos que operan en coordinación con redes nacionales o europeas de datos espaciales. Su implementación de ISO 19115 suele estar integrada en sistemas más amplios de gestión geoespacial, como Infraestructuras de Datos Espaciales (IDE), lo que asegura una alineación con estándares internacionales.

Aunque su número es reducido, estos repositorios presentan un alto nivel de sofisticación técnica y documental, y sus metadatos incluyen información detallada sobre calidad, procesamiento y validación, lo cual refuerza significativamente la verificación de autenticidad.

ISO 19115 incorpora múltiples mecanismos para garantizar la autenticidad de los datos:

- Documentación estructurada del linaje técnico e institucional
- Control de calidad a través de evaluaciones normalizadas

- Identificación de agentes y fuentes en cada etapa del ciclo de vida
- Integración con sistemas de preservación digital y auditoría

Si bien no incluye de forma nativa elementos como *checksums* o firmas digitales, estos pueden ser incorporados a través de extensiones o en combinación con otros estándares complementarios (ej. METS, PREMIS).

Una ventaja estratégica de ISO 19115 es su alta interoperabilidad con otros modelos, permitiendo su integración en infraestructuras científicas multidisciplinares. Su adopción facilita el cumplimiento de los principios FAIR, particularmente en lo relativo a la trazabilidad y la reutilización de los datos, siempre que se implementen mecanismos técnicos adicionales de verificación.

En síntesis, ISO 19115 representa una solución robusta y especializada para describir y verificar la autenticidad de datos geospaciales, siendo particularmente adecuada para repositorios con funciones cartográficas, ambientales o territoriales.

#### 4.6. ANÁLISIS COMPARATIVO DE MODELOS

Para evaluar de manera sistemática las capacidades de los cinco modelos de metadatos analizados (Dublin Core, DataCite, OAI-ORE, DDI e ISO 19115) en relación con la verificación de autenticidad, se construyó una matriz comparativa basada en tres dimensiones fundamentales: procedencia, integridad y trazabilidad. A cada modelo se le asignó una puntuación cualitativa (alta, media o baja) según su estructura formal y las prácticas observadas en su implementación por los repositorios españoles.

Tabla 1. Matriz comparativa

Modelo	Procedencia	Integridad técnica	Trazabilidad / versionado
Dublin Core	Baja	Baja	Baja
DataCite	Alta	Media-alta	Alta
OAI-ORE	Media	Baja	Media-alta
DDI	Alta	Media	Alta
ISO 19115	Alta	Alta	Alta

La Tabla 1 revela que los esquemas más robustos en autenticidad son aquellos especializados y diseñados para contextos específicos, como ISO 19115 en datos geográficos, DDI en ciencias sociales y DataCite en datos científicos en general. En contraste, Dublin Core, pese a su amplia adopción, presenta debilidades estructurales en los tres aspectos evaluados.

Un análisis DAFO (debilidades, amenazas, fortalezas y oportunidades) puede resumir cada uno de los modelos, desde la perspectiva de la autenticidad:

- Dublin Core
  - Debilidades: escasa capacidad para documentar versiones y agentes; falta de soporte para verificación técnica.
  - Fortalezas: amplia adopción, interoperabilidad básica.
  - Amenazas: estancamiento normativo; dependencia de perfiles ad hoc.
  - Oportunidades: desarrollo de extensiones interoperables; mapeo con modelos más ricos.
- DataCite
  - Debilidades: requiere integración adicional para preservación técnica.
  - Fortalezas: identificación persistente; buena documentación de procedencia.
  - Amenazas: subutilización en repositorios de tamaño pequeño.
  - Oportunidades: alineación con FAIR; interoperabilidad con ORCID, Crossref y OpenAIRE.
- OAI-ORE
  - Debilidades: falta de estandarización semántica; escaso soporte documental en español.
  - Fortalezas: representación de agregaciones complejas.
  - Amenazas: alta complejidad técnica.
  - Oportunidades: uso combinado con otros modelos; trazabilidad relacional avanzada.
- DDI
  - Debilidades: especializado; curva de aprendizaje elevada.
  - Fortalezas: ciclo de vida completo de los datos; contexto metodológico.
  - Amenazas: implementación limitada fuera de ciencias sociales.
  - Oportunidades: interoperabilidad con estándares estadísticos; uso creciente en estudios longitudinales.
- ISO 19115
  - Debilidades: no nativo en ambientes no geoespaciales.
  - Fortalezas: documentación exhaustiva del linaje y calidad.
  - Amenazas: bajo número de implementaciones en repositorios generales.
  - Oportunidades: integración con IDE; cumplimiento con políticas europeas (INSPIRE).

El análisis sugiere que no existe un modelo «universalmente superior», sino que la selección debe responder a las características del dominio disciplinar, la complejidad de los datos y la infraestructura tecnológica del repositorio. Por ejemplo:

- *DataCite* resulta adecuado para repositorios interdisciplinarios que priorizan la citación y trazabilidad de datos.
- *DDI* es idóneo para archivos de datos de encuestas y estudios sociales.

- *ISO 19115* se recomienda para información geoespacial que requiere validación precisa del linaje.
- *OAI-ORE* es útil en repositorios que gestionan objetos digitales compuestos o agregados.
- *Dublin Core* puede mantenerse como capa básica interoperable, pero debería complementarse con otros esquemas más específicos.

El patrón de adopción observado en los repositorios españoles refleja una tensión entre facilidad de implementación y sofisticación funcional. Mientras que Dublin Core domina en número, su limitada capacidad técnica plantea desafíos para la calidad y la autenticidad de los datos preservados. En contraste, los esquemas más avanzados (como DataCite, DDI o ISO 19115) tienden a ser utilizados por repositorios con mayor madurez técnica o enfoque temático específico.

Asimismo, se observan tendencias emergentes de hibridación, en las que un modelo básico (como Dublin Core) se utiliza junto a modelos más ricos, mediante extensiones, mapeos o perfiles de aplicación locales. Esta estrategia permite mejorar la capacidad de verificación sin abandonar la interoperabilidad mínima.

Por último, la evolución normativa internacional –particularmente, en torno a los principios FAIR y TRUST– está generando incentivos para que los repositorios revisen sus modelos de metadatos y adopten soluciones que permitan una verificación estructurada de la autenticidad, como requisito para la certificación y el financiamiento público.

Por lo tanto, se puede concluir que Dublin Core, pese a ser el modelo más utilizado (33 repositorios), ofrece una estructura limitada para documentar procedencia, verificar integridad o registrar cambios en los datos. Su implementación en la mayoría de los casos se reduce a los elementos básicos, sin cualificadores ni extensiones orientadas a la preservación digital. La generalización de este modelo parece obedecer más a su facilidad técnica e integración con plataformas comunes (como DSpace) que a una decisión basada en criterios de autenticidad documental.

En contraste, DataCite (7 repositorios), DDI (3) e ISO 19115 (2) demuestran un mayor grado de especialización y robustez estructural para documentar aspectos críticos de la autenticidad. DataCite destaca por su integración con identificadores persistentes y su capacidad para representar agentes y versiones; DDI sobresale por su exhaustividad en describir procesos metodológicos y transformaciones de datos; e ISO 19115 aporta mecanismos precisos para rastrear el linaje y la calidad de datos geográficos.

OAI-ORE, por su parte, ofrece un enfoque distintivo orientado a representar agregaciones complejas, lo que resulta útil en contextos donde los objetos digitales están distribuidos y organizados de forma modular. No obstante, su implementación en repositorios españoles es limitada y técnicamente heterogénea.

Los resultados sugieren varias oportunidades concretas para fortalecer la capacidad del sistema español en materia de autenticidad de datos de investigación:

- Revisión y ampliación de los perfiles de metadatos existentes, especialmente en repositorios que utilizan Dublin Core, incorporando calificadores, elementos de preservación y mecanismos de verificación técnica.
- Fomento de la adopción de modelos especializados, como DataCite, DDI o ISO 19115, mediante capacitación técnica, incentivos institucionales y marcos de apoyo normativo.
- Desarrollo de perfiles de aplicación interoperables que integren múltiples esquemas en una arquitectura coherente, permitiendo mantener la simplicidad en el acceso y la interoperabilidad básica, sin renunciar a una descripción documental robusta.

Asimismo, podría promoverse el uso combinado de OAI-ORE y esquemas de preservación (como PREMIS), particularmente en repositorios que gestionan objetos digitales compuestos, versiones múltiples o conjuntos distribuidos.

## 5. CONCLUSIONES

Este estudio ha ofrecido una evaluación de cinco modelos de metadatos implementados en repositorios españoles de datos de investigación, con especial atención a su capacidad para garantizar la autenticidad digital. A partir del análisis de 63 repositorios registrados en RE3DATA, se identificaron fortalezas y limitaciones estructurales en los modelos Dublin Core, DataCite, OAI-ORE, DDI e ISO 19115.

Los resultados demuestran una brecha entre la adopción de modelos de metadatos y la efectiva implementación de mecanismos que garantizan la autenticidad de los datos en repositorios. Mientras Dublin Core continúa siendo el esquema predominante por su simplicidad, los modelos especializados como DataCite, DDI e ISO 19115 ofrecen soluciones más robustas para preservar la autenticidad, condición indispensable para la reproducibilidad científica y la confianza en los datos.

La investigación también destaca el papel estratégico de RE3DATA como registro global y herramienta de monitoreo del ecosistema de repositorios, así como la necesidad de alinear las prácticas nacionales con las directrices internacionales (FAIR, TRUST, COAR, OpenAIRE).

Estas observaciones plantean importantes consideraciones para el desarrollo futuro de repositorios de datos de investigación en España. La adopción de modelos de metadatos que incorporen elementos robustos para la verificación de autenticidad resulta fundamental para garantizar la confiabilidad y la reutilización de los datos científicos a largo plazo. Se recomienda la implementación de extensiones específicas o bien la adopción de esquemas complementarios que subsanen las limitaciones identificadas, particularmente en repositorios que utilizan Dublin Core como esquema principal.

Adicionalmente, se sugiere la creación de directrices específicas a nivel nacional, que establezcan requisitos mínimos para la documentación de autenticidad en los metadatos de datos de investigación, alineados con estándares internacionales como los Principios FAIR y las recomendaciones de COAR sobre buenas prácticas para repositorios.

Para concluir, este estudio contribuye al campo emergente de la curación y la preservación de datos de investigación, proporcionando una evaluación sistemática de las capacidades de los modelos de metadatos actuales para garantizar un aspecto crítico de la preservación digital: la autenticidad. Los resultados obtenidos pueden informar tanto a gestores de repositorios como a responsables de políticas científicas sobre las fortalezas y las debilidades de los diferentes esquemas de metadatos en relación con la verificación de autenticidad, facilitando decisiones informadas en la selección y la implementación de estos esquemas en futuros desarrollos.

La autenticidad de los datos de investigación constituye un pilar necesario para una ciencia abierta rigurosa, responsable y socialmente confiable. Garantizar esta autenticidad requiere no solo tecnología, sino también gobernanza, formación y compromiso institucional. En este sentido, los modelos de metadatos no son simples herramientas descriptivas, sino instrumentos de garantía epistemológica y transparencia científica.

## REFERENCIAS BIBLIOGRÁFICAS

- Abadal Falgueras, E. y Anglada Ferrer, L. (2020). Ciencia abierta: cómo han evolucionado la denominación y el concepto. *Anales de Documentación*, 23(1), 1. <https://doi.org/10.6018/analesdoc.378171>
- COAR. (2022). *Community Framework for Good Practices in Repositories*, version 2. Zenodo. <https://zenodo.org/records/7108101>.
- Cunha, P., Marks Löw, M., Santos, A. A. dos y Cândido, A. C. (2024). Competência informacional para uso de Dados abertos: tendências e perspectivas. *Em Questao*, 30. <https://doi.org/10.1590/1808-5245.30.136545>
- DataCite. *Annual Report*. (2021). [https://datacite.org/wp-content/uploads/2023/06/DataCite\\_2021\\_Annual\\_Report.pdf](https://datacite.org/wp-content/uploads/2023/06/DataCite_2021_Annual_Report.pdf)
- DataCite Metadata Working Group. (2024). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs*. Version 4.6. DataCite e.V. <https://doi.org/10.14454/mzv1-5b55>
- De Giusti, M. R. (2021). Calidad en los repositorios digitales. Los principios TRUST para repositorio de datos. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 29, 55-59. <https://doi.org/10.24215/18509959.29.e6>
- DublinCore. (2025). *DublinCore Specifications*. <https://www.dublincore.org/specifications/dublin-core/>
- Duranti, L. (2007). The InterPARES 2 Project (2002-2007): An Overview. *Archivaria*, 64, 113-121. <https://www.archivaria.ca/index.php/archivaria/article/view/13155>
- Duranti, L., Rogers, C. y Thibodeau, K. (2022). Authenticity. *Archives and Records*, 43(2), 188-203. <https://doi.org/10.1080/23257962.2022.2054406>
- Fundación Española para la Ciencia y la Tecnología (FECYT). (2021). *Guía para la evaluación de repositorios institucionales de investigación v4.0*. <https://www.fecyt.es/publicaciones/guia-para-la-evaluacion-de-repositorios-institucionales-de-investigacion>
- Gallego Lorenzo, J., Martín González, Y. e Iglesias Rodríguez, A. (2024). Calidad y uso de datos confiables en la práctica docente universitaria. *Ibersid: Revista de Sistemas de Información y Documentación*, 18(2), 87-96. <https://doi.org/10.54886/ibersid.v18i2.5019>
- Jurado Martín, M. (2023). Competencia crítica y desinformación en teoría de la comunicación: Análisis en los programas de los grados en periodismo en España. *Documentación de las Ciencias de la Información*, 46(2), 169-176. <https://doi.org/10.5209/dcin.87513>

- Kanza, S. y Knight, N. J. (2022). Behind every great research project is great data management. *BMC Research Notes*, 15(20). <https://doi.org/10.1186/s13104-022-05908-5>
- Khan, A. M., Loan, F. A., Parray, U. Y. y Rashid, S. (2024). Global overview of research data repositories: an analysis of re3data registry. *Information Discovery and Delivery*, 52(1), 53-61. <https://doi.org/10.1108/IDD-07-2022-0069>
- Lagoze, C., Van De Sompel, H., Nelson, M. L., Warner, S., Sanderson, R. y Johnston, P. (2008). *Object re-use and exchange: A resource-centric approach*. Indianapolis. <http://arxiv.org/abs/0804.2273>
- Lin, D., Crabtree, J., Dillo, I., Downs, R., Edmunds, R., Giarretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stochhause, M. y Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(144). <https://www.nature.com/articles/s41597-020-0486-7>
- Martín González, Y. e Iglesias Rodríguez, A. (2023). Alfabetización en datos: Prácticas y escenarios formativos. *Revista EDICIC*, 3(3), 1-15. <https://ojs.edicic.org/revistaedicic/article/view/202>
- OpenAIRE. (2022). *OpenAIRE Guidelines*. <https://guidelines.openaire.eu/en/latest/index.html>
- Pacheco, A., Da Silva, C. G. y De Freitas, M. C. V. (2023). A metadata model for authenticity in digital archival descriptions. *Archival Science*, 23, 629-673. <https://doi.org/10.1007/s10502-023-09422-w>
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirmbacher, P. y Dierolf, U. (2013) Making Research Data Repositories Visible: The re3data.org Registry. *PLoS ONE*, 8(11), e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Silva, V. de A. y Silva, M. B. da. (2021). Metadatos para la preservación digital de datos abiertos: Un estudio de identificación. *Biblios Journal of Librarianship and Information Science*, 78, 44-60. <https://doi.org/10.5195/biblios.2020.793>
- Vardigan, M. y Whiteman, C. (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. *Archival Science*, 7, 73-87. <https://doi.org/10.1007/s10502-006-9037-z>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I.-J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Da-Silva-Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, Chris T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

## UN MARCO DE MÉTRICAS SMART PARA LA MEJORA DE LOS PLANES DE GESTIÓN DE DATOS

### *A SMART METRICS FRAMEWORK FOR IMPROVING DATA MANAGEMENT PLANS*

Federico GRASSO TORO

*Universidad de Berna, Suiza*

Federico.Grasso@unibe.ch

 <https://orcid.org/0000-0002-9041-0868>

**RESUMEN:** Este artículo propone un marco innovador para mejorar los planes de gestión de datos (PGD), transformándolos en herramientas integrales para administrar diversos activos intelectuales más allá de los datos de investigación tradicionales, incluyendo software y modelos. El marco integra cuatro metodologías clave: métricas SMART (Específicas, Medibles, Alcanzables, Relevantes, con Plazos Determinados) para el establecimiento y el seguimiento claro de objetivos; la gestión del ciclo de vida de los datos de investigación (RDLC) para un enfoque holístico desde la concepción hasta la reutilización de los datos; los principios FAIR (Encontrables, Accesibles, Interoperables y Reutilizables) para maximizar el valor y el impacto de los datos; y el desarrollo dirigido por el comportamiento (BDD) para alinear las funcionalidades de los PGD con las necesidades de los usuarios y mejorar la comunicación. Este marco integral busca mejorar la investigación optimizando la planificación, la ejecución y la evaluación, promoviendo la reproducibilidad y la transparencia. Facilita la creación de PGD claros y automatizados, aplicándose actualmente en ensayos clínicos, ciencias ambientales y humanidades digitales. Se espera que mejore la comunicación entre máquinas, automatice tareas y aumente la eficiencia y la transparencia en la investigación.

**PALABRAS CLAVE:** planes de gestión de datos; métricas SMART; ciclo de vida de los datos de investigación; principios FAIR; desarrollo dirigido por el comportamiento; sintaxis Gherkin.

**ABSTRACT:** This paper proposes an innovative framework to enhance Data Management Plans (DMPs), transforming them into comprehensive tools for managing diverse intellectual assets beyond traditional research data, including software and models. The framework integrates four key methodologies: SMART metrics (Specific, Measurable, Achievable, Relevant, Time-bound) for clear objective setting and monitoring; Research Data Life Cycle (RDLC) management for a holistic approach from data conception to reuse; FAIR principles

(Findable, Accessible, Interoperable, Reusable) to maximize data value and impact; and Behavior-Driven Development (BDD) to align DMP functionalities with user needs and improve communication. This holistic framework optimizes research planning, execution, and evaluation for reproducibility and transparency. It creates machine-readable, actionable Data Management Plans (DMPs) through clear scenarios and automated validation. Initial applications in clinical trials, environmental sciences, and digital humanities demonstrate its potential to improve data quality, streamline processes, and enhance collaboration. The framework aims to enhance machine-to-machine communication, automate DMP tasks, and increase research efficiency and transparency.

KEYWORDS: data management plans; SMART metrics; research data life cycle; FAIR principles; behavior-driven development; Gherkin syntax.

## 1. INTRODUCCIÓN

En el panorama científico actual, los planes de gestión de datos han trascendido su función tradicional de recopilar, almacenar y compartir datos (Dataversity, s. f.). Ahora se pueden visualizar como herramientas integrales para gestionar una gama más amplia de activos intelectuales, incluyendo software, modelos, datos y otros resultados de investigación.

### 1.1. HACIA UN NUEVO MARCO PARA LOS PLANES DE GESTIÓN DE DATOS

La evolución de la ciencia abierta exige un marco más robusto y transparente que facilite la comunicación entre investigadores, instituciones académicas y entidades de financiación. Este artículo propone un marco innovador que integra (1) las métricas SMART (de sus siglas en inglés, traducible a Específico, Medible, Alcanzable, Relevante y con Plazos Determinados de Tiempo); (2) la gestión del ciclo de vida de los datos de investigación (RDLC, de sus siglas en inglés); (3) los principios FAIR (de la siglas en inglés, traducible a Encontrable, Accesible, Interoperable y Reutilizable), y (4) el desarrollo dirigido por el comportamiento (BDD, de sus siglas en inglés) para mejorar el manejo de los PGD. Este marco holístico no solamente busca optimizar la planificación, ejecución y evaluación de las actividades de investigación, sino también promover la reproducibilidad y la transparencia en la ciencia.

### 1.2. HACIA UN MANEJO EXTENDIDO DE LOS PLANES DE GESTIÓN DE DATOS

La dinámica y creciente complejidad de la investigación científica contemporánea demanda imperativamente la adopción de estrategias de gestión de datos que trascienden los enfoques tradicionales (Burnette *et al.*, 2016). La proliferación de datos, su diversidad tipológica y la necesidad de colaboración a gran escala entre actores dispares –investigadores de diversas disciplinas, instituciones académicas con diferentes enfoques y entidades de financiación con

requisitos específicos— configuran un escenario donde la eficiencia, la transparencia y la reproducibilidad se erigen como pilares fundamentales de la ciencia actual.

En este contexto, la mera planificación de la gestión de datos a través de los Planes de Gestión de Datos (PGD) se revela como un eslabón crucial, pero intrínsecamente limitado si no se sustenta en un andamiaje metodológico robusto y adaptable a activos intelectuales que van más allá de datos de investigación. Este marco se concibe como una amalgama sinérgica de metodologías y principios de probada eficacia en diversos dominios, integrados específicamente para potenciar la planificación, la ejecución, la evaluación y, en última instancia, el impacto de la investigación científica (Grasso Toro, 2024). Los componentes centrales son:

1. La integración de métricas SMART: La definición de objetivos de investigación y la planificación de la gestión de datos deben regirse por los criterios SMART [4] (Específico, Medible, Alcanzable, Relevante y con Plazos Determinados de Tiempo). Esta integración asegura que cada aspecto del PGD esté claramente definido, pueda ser monitorizado a través de indicadores concretos, sea realista en función de los recursos disponibles y las capacidades del equipo, contribuya de manera significativa a los objetivos generales de la investigación y esté delimitado temporalmente. La aplicación sistemática de métricas SMART a la gestión de datos facilita la evaluación del progreso, la identificación temprana de posibles desviaciones y la toma de decisiones informadas para la optimización del proceso investigativo.
2. La gestión del ciclo de vida de los datos de investigación (RDLC): Un enfoque integral de la gestión de datos requiere una perspectiva que abarque la totalidad del ciclo de vida de los datos, desde su concepción y recopilación hasta su preservación a largo plazo y potencial reutilización. El RDLC proporciona una estructura para identificar las diferentes etapas por las que transitan los datos de investigación —planificación, captura, procesamiento, análisis, preservación y reutilización— y para definir las acciones y las responsabilidades correspondientes en cada una de ellas. La adopción consciente del RDLC en la elaboración de los PGD garantiza una visión holística de la gestión de datos, anticipando las necesidades en cada fase y promoviendo prácticas consistentes y eficientes.
3. La adherencia a los principios FAIR: Los principios FAIR (Encontrable, Accesible, Interoperable y Reutilizable) constituyen un conjunto de directrices esenciales para maximizar el valor y el impacto de los datos de investigación. La integración de estos principios en el diseño y la implementación de los PGD asegura que los datos generados sean fácilmente descubiertos por otros investigadores (Encontrable), que puedan ser obtenidos y procesados bajo condiciones claras (Accesible), que puedan combinarse, compararse e intercambiarse con otros conjuntos de datos (Interoperable) y que puedan ser utilizados para futuras investigaciones con la atribución adecuada (Reutilizable). La adopción de los principios FAIR fomenta la ciencia abierta, la colaboración y la acumulación valiosa de conocimiento científico.
4. La implementación del desarrollo dirigido por el comportamiento (BDD): El desarrollo dirigido por el comportamiento (BDD) es una metodología de desarrollo de software que se centra en la especificación del comportamiento del sistema desde la

perspectiva del usuario. Para la gestión de PGD, la adaptación de los principios del BDD puede aportar una valiosa perspectiva en la definición de los requisitos y las expectativas en relación con la gestión de datos. Al describir el comportamiento deseado del sistema de gestión de datos en términos de escenarios concretos y resultados observables, se facilita la comunicación entre los diferentes actores involucrados (i. e., investigadores, gestores de datos, responsables de infraestructura y entidades de financiación) y se asegura que el PGD resultante responda de manera efectiva a las necesidades de las actividades de investigación. El BDD ayuda a traducir estas con mayor claridad a procesos de gestión de datos, determinando los roles y responsabilidades, además de los criterios de éxito para su evaluación formal.

La convergencia de estas cuatro dimensiones –métricas SMART, gestión del ciclo de vida de los datos, principios FAIR y desarrollo dirigido por el comportamiento– da como resultado un marco holístico para la gestión avanzada de los PGD. Este marco no se limita a ser una herramienta de planificación estática, sino que se concibe como un enfoque dinámico y adaptable que guía la ejecución de la investigación, facilita la evaluación continua del proceso de gestión de datos y promueve una cultura de transparencia y reproducibilidad en la ciencia.

Al adoptar este marco integral, las instituciones académicas, los investigadores y las entidades de financiación pueden optimizar sus inversiones en investigación, maximizar el impacto de los descubrimientos científicos para avanzar hacia un ecosistema de investigación más abierto, colaborativo y eficiente.

## 2. METODOLOGÍA

### 2.1. MÉTRICAS SMART: UN PILAR FUNDAMENTAL PARA LA PLANIFICACIÓN Y LA EVALUACIÓN DE LA INVESTIGACIÓN

Las métricas SMART (Corportate Finance Institute, s. f.) proporcionan una base sólida para definir objetivos de investigación claros, medibles y alcanzables. Al aplicar este enfoque a los PGD, se facilita la alineación de los paquetes de trabajo y los entregables con los objetivos generales del proyecto.

Esto permite una mejor gestión de los recursos y una evaluación más precisa del progreso de la investigación. La integración de las métricas SMART en los PGD puede ofrecer la siguiente serie de ventajas:

- Claridad y precisión: Se definen objetivos específicos y medibles, evitando ambigüedades y facilitando la evaluación.
- Alineación estratégica: Se asegura que los esfuerzos de investigación estén alineados con los objetivos generales del proyecto.
- Evaluación objetiva: Se establecen criterios medibles para evaluar progreso y éxito.
- Rendición de cuentas: Se promueve la transparencia y la responsabilidad en el uso de los recursos.

### 2.1.1. *Ventajas de un enfoque estructurado en la investigación*

Un enfoque estructurado en la investigación se fundamenta en la definición clara y precisa de los objetivos que se persiguen. Esto no solo evita confusiones y desviaciones durante el proceso investigativo, sino que también facilita enormemente la posterior evaluación de los resultados obtenidos.

Al establecer objetivos específicos y, crucialmente, medibles, se crea un marco de referencia tangible contra el cual se puede contrastar el progreso y determinar el nivel de éxito alcanzado. La ausencia de ambigüedades en la formulación de los objetivos es, por tanto, un pilar fundamental de este enfoque.

En segundo lugar, un enfoque estructurado garantiza una sólida alineación estratégica entre las actividades de investigación y los objetivos generales del proyecto en el que se enmarcan.

Esta coherencia es vital para asegurar que los recursos invertidos en la investigación contribuyan de manera efectiva a la consecución de las metas superiores. Al mantener los esfuerzos de investigación intrínsecamente ligados a la estrategia global, se maximiza su impacto y se evita la dispersión de recursos en líneas de investigación periféricas o irrelevantes para el propósito principal del proyecto.

La evaluación objetiva constituye otro beneficio esencial de este enfoque. Al establecer criterios de evaluación que son inherentemente medibles, se facilita un análisis imparcial y basado en datos concretos del progreso realizado y del éxito final de la investigación.

Estos criterios predefinidos permiten superar las valoraciones subjetivas y proporcionan una base sólida para la toma de decisiones informadas a lo largo del ciclo de vida del proyecto. La capacidad de medir objetivamente los avances es fundamental para identificar fortalezas, debilidades y áreas de mejora.

Finalmente, un enfoque estructurado en la investigación fomenta de manera inherente la rendición de cuentas. La definición clara de objetivos, la alineación estratégica con las metas del proyecto y la implementación de criterios de evaluación medibles contribuyen a una mayor transparencia en el uso de los recursos asignados a la investigación.

Asimismo, promueven la responsabilidad individual y colectiva en la ejecución de las tareas y en la consecución de los resultados esperados. Esta transparencia y responsabilidad son cruciales para la gestión eficiente de proyectos de investigación y para la optimización del uso de los recursos disponibles.

## 2.2. GESTIÓN DEL CICLO DE VIDA DE LOS DATOS DE INVESTIGACIÓN (RDLC): PRESERVANDO LA INTEGRIDAD Y EL VALOR DE LOS DATOS

El RDLC proporciona un enfoque estructurado para gestionar los datos de investigación a lo largo de todo su ciclo de vida, desde su creación hasta su disposición final (Schmitt *et al.*, s. f.).

Al integrar los principios FAIR (Wilkinson *et al.*, 2016) en cada etapa del RDLC, se garantiza que los datos sean fáciles de encontrar, acceder, interoperar y reutilizar, ofreciendo los beneficios de:

- Preservación de la calidad de los datos: Se implementan medidas para asegurar la integridad y la calidad de los datos a lo largo del tiempo.
- Mejora de la accesibilidad: Se facilita el acceso a los datos por parte de otros investigadores y la comunidad en general.
- Fomento de la reutilización: Se promueve la reutilización de los datos para nuevas investigaciones y aplicaciones.
- Cumplimiento de las normativas: Se facilita el cumplimiento de las normativas y políticas de gestión de datos.

### 2.2.1. Beneficios de una gestión responsable de datos de investigación

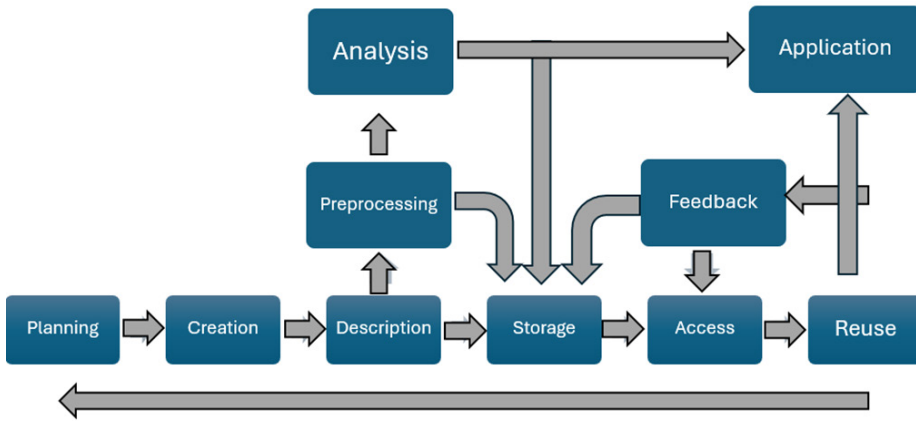
Una gestión responsable de los datos de investigación aporta múltiples beneficios cruciales para la comunidad científica y la sociedad en general:

- *Preservación a largo plazo de la calidad de los datos*: Se establecen y ejecutan protocolos meticulosos para salvaguardar la exactitud, la coherencia y la fiabilidad de los datos durante todo su ciclo de vida. Esto incluye la implementación de técnicas de documentación exhaustivas, formatos de archivo duraderos, validación de datos y estrategias de almacenamiento seguro para prevenir la corrupción o la pérdida de información valiosa con el paso del tiempo. La preservación de la calidad garantiza que las conclusiones de la investigación sigan siendo válidas y puedan ser verificadas en el futuro.
- *Optimización de la accesibilidad para la comunidad investigadora y el público*: Se desarrollan e implementan mecanismos y plataformas que simplifican y amplían el acceso a los conjuntos de datos generados por la investigación. Esto implica la utilización de metadatos descriptivos y estandarizados, la creación de repositorios de datos en línea de fácil navegación, la aplicación de licencias de uso claras y la adopción de estándares de interoperabilidad. Un acceso facilitado promueve la colaboración científica, evita la duplicación de esfuerzos y permite que un espectro más amplio de investigadores y el público interesado puedan analizar y utilizar los datos para avanzar en el conocimiento y la innovación.
- *Impulso a la reutilización de datos para nuevas investigaciones y aplicaciones innovadoras*: Se fomenta activamente la exploración y el empleo de datos existentes para abordar nuevas preguntas de investigación o para desarrollar aplicaciones novedosas en diversos campos. Esto se logra mediante la promoción de una cultura de intercambio de datos; la provisión de herramientas y recursos para facilitar el descubrimiento y la comprensión de los datos, y el reconocimiento del valor de la reutilización como un motor de eficiencia y progreso científico. La reutilización maximiza el retorno de la inversión en investigación y acelera el ritmo de los descubrimientos.
- *Facilitación del cumplimiento de normativas y políticas de gestión de datos*: Se implementan estrategias y procedimientos que aseguran la adhesión a las directrices éticas, legales y regulatorias relacionadas con la recopilación, el almacenamiento, el uso y la

difusión de datos de investigación. Esto incluye la consideración de aspectos como la protección de la privacidad, la confidencialidad de los datos sensibles, los derechos de propiedad intelectual y las políticas institucionales o de los organismos financiadores. Un enfoque proactivo en el cumplimiento normativo genera confianza en la integridad de la investigación y evita posibles sanciones o problemas legales.

La Figura 1 presenta el ciclo de vida de los datos de investigación, pensando su reutilización como objetivo adicional a su aplicación inicial.

Figura 1. Gestión del ciclo de vida de los datos de investigación (RDLC) (Schmitt *et al.*, s. f.)



### 2.3. PRINCIPIOS FAIR: FACILITANDO EL DESCUBRIMIENTO, EL ACCESO Y LA REUTILIZACIÓN DE LOS DATOS

Los principios FAIR (Wilkinson *et al.*, 2016) son esenciales para maximizar el valor de los datos de investigación. Al aplicar estos principios, se asegura que los datos sean, descritos en sus traducciones al español, fáciles de:

- Encontrar: Fáciles de encontrar por parte de humanos y máquinas.
- Acceder: Accesibles de forma sencilla y con las restricciones necesarias.
- Interoperar: Compatibles con otros sistemas y conjuntos de datos.
- Reutilizar: Adecuadamente documentados para facilitar su reutilización.

La implementación de los principios FAIR no solo beneficia a los investigadores, sino también a las instituciones académicas, las entidades financiadoras y la sociedad en general. La implementación de los principios FAIR (es decir, que los datos sean Encontrables, Accesibles, Interoperables y Reutilizables) en la investigación no solo beneficia a los investigadores individuales al facilitar la gestión y el intercambio de datos, sino que también tiene un impacto positivo más amplio. La adopción de los principios FAIR representa una estrategia de

gran alcance con efectos positivos que trascienden la labor individual de los investigadores, extendiéndose de manera significativa a las instituciones académicas en su conjunto, a las organizaciones que proporcionan el soporte financiero para la investigación y, en última instancia, a la sociedad en su totalidad. La internalización y la puesta en práctica de estos principios fundamentales —que dictaminan que los datos generados en el curso de la investigación deben ser Encontrables mediante metadatos ricos y persistentes, Accesibles bajo condiciones claras y protocolos estandarizados, Interoperables a través de formatos comunes y vocabularios controlados y Reutilizables con licencias explícitas y documentación detallada sobre su procedencia y uso— no se limitan a simplificar la administración y el intercambio de la información científica para los investigadores. Su influencia se proyecta en una mejora sustancial de la eficiencia y la transparencia del proceso científico en general.

Para las instituciones académicas, la adopción de los principios FAIR fomenta una cultura de ciencia abierta y colaboración, incrementando la visibilidad y el impacto de su investigación. Al facilitar el descubrimiento y la reutilización de los datos producidos por sus investigadores, las instituciones pueden optimizar la inversión en investigación, evitar la duplicación de esfuerzos y promover la innovación a partir de trabajos previos.

Asimismo, la gestión FAIR de los datos fortalece la reputación de la institución al demostrar un compromiso con las mejores prácticas en la investigación y la transparencia en la difusión de sus resultados.

Las entidades financiadoras, por su parte, se benefician de la implementación de los principios FAIR al asegurar un mayor retorno de su inversión. La disponibilidad y la reutilización de los datos financiados permiten maximizar el impacto de la investigación, facilitando la validación de resultados, la generación de nuevas hipótesis y el avance del conocimiento en diversas áreas. Además, la transparencia inherente a los datos FAIR puede mejorar la rendición de cuentas y la evaluación del impacto de los proyectos financiados.

Finalmente, la sociedad en general se ve favorecida por la adopción de los principios FAIR a través de un acceso más amplio y eficiente al conocimiento científico. La posibilidad de encontrar, acceder, interoperar y reutilizar datos de investigación facilita la innovación; la toma de decisiones basada en evidencia, y el desarrollo de soluciones a los desafíos sociales, económicos y ambientales. En última instancia, la ciencia abierta y los datos FAIR contribuyen a una sociedad más informada, participativa y capaz de abordar los retos del futuro.

#### 2.4. DESARROLLO DIRIGIDO POR EL COMPORTAMIENTO (BDD): CERRANDO LA BRECHA HUMANOS-MÁQUINAS

El BDD es una metodología que facilita la creación y la evaluación de escenarios de investigación, asegurando que las métricas SMART sean claras, accionables y comprobables.

Al utilizar un lenguaje natural y una sintaxis estructurada (llamada Gherkin), el BDD promueve la colaboración entre investigadores, gestores de datos y otros actores involucrados en el proceso de investigación.

El BDD ofrece las siguientes ventajas:

- **Mejora de la comunicación:** Se facilita la comunicación entre los diferentes actores involucrados en la investigación.
- **Claridad y precisión:** Se definen escenarios de investigación claros y precisos, evitando ambigüedades.
- **Validación continua:** Se promueve la validación continua de las métricas SMART a través de pruebas automatizadas.
- **Trazabilidad y documentación:** Se facilita el seguimiento de los cambios y la documentación del proceso de investigación.

#### 2.4.1. *Beneficios de la metodología propuesta*

La implementación de la metodología propuesta para la gestión de coloquios de investigación aporta una serie de ventajas significativas que impactan positivamente en el desarrollo y la calidad del proceso investigativo:

- *Mejora sustancial de la comunicación:* Se establece un marco que facilita y estructura la comunicación entre todos los actores involucrados en la investigación, incluyendo investigadores principales, colaboradores, mentores y revisores. Esto se traduce en un flujo de información más eficiente, la reducción de malentendidos y una mayor colaboración en la resolución de desafíos. La comunicación clara y abierta fomenta un ambiente de trabajo más productivo y cohesivo.
- *Establecimiento de claridad y precisión en los escenarios de investigación:* La metodología promueve la definición de escenarios de investigación con un alto grado de claridad y precisión. Esto implica la delimitación exhaustiva de los objetivos, el alcance, las preguntas de investigación y la metodología a seguir. Al evitar ambigüedades desde el inicio, se minimizan las interpretaciones erróneas y se asegura que todos los participantes compartan una comprensión común del proyecto, lo que contribuye a una ejecución más enfocada y eficiente.
- *Fomento de la validación continua mediante pruebas automatizadas:* Un aspecto crucial de la metodología es la integración de la validación continua de las métricas SMART (Específicas, Medibles, Alcanzables, Relevantes y con Plazo definido) a través de la implementación de pruebas automatizadas. Esta práctica asegura que el progreso de la investigación se evalúe de manera objetiva y regular, permitiendo la detección temprana de posibles desviaciones o problemas. La automatización de las pruebas ahorra tiempo y recursos, al tiempo que aumenta la fiabilidad de la validación.
- *Optimización de la trazabilidad y la documentación del proceso de investigación:* La metodología facilita un seguimiento detallado de todos los cambios, decisiones y resultados que surgen a lo largo del proceso de investigación. Asimismo, promueve la documentación exhaustiva de cada etapa, desde la concepción inicial hasta la presentación de los resultados. Esta trazabilidad y esta documentación detallada no solo facilitan la revisión y el análisis del progreso, sino que también son fundamentales para la

replicabilidad de la investigación, la transferencia de conocimiento y el cumplimiento de los estándares académicos y científicos.

## 2.5. APLICACIÓN PRÁCTICA DEL MARCO: UN EJEMPLO DE IMPLEMENTACIÓN

Para ilustrar la aplicación práctica del marco, se presenta un ejemplo de cómo el BDD puede utilizarse para crear métricas SMART para PGD «legibles por máquina» y potencialmente «accionables por máquina» en una futura instancia.

El proceso se divide en cuatro pasos:

- Descubrir y redactar escenarios de investigación claros: Se identifican los comportamientos clave de los científicos y se traducen en escenarios concretos y detallados utilizando un lenguaje claro y sin ambigüedades (mediante sintaxis Gherkin).
- Formular métricas a partir de estos escenarios: Se definen resultados medibles relativos a las métricas SMART para cada escenario con el fin de evaluar objetivamente el éxito por evaluar.
- Automatizar pruebas para escenarios: Se transforman los escenarios en pruebas automatizadas que permiten la validación continua del sistema en relación con los objetivos definidos e hitos.
- Supervisar y repetir las descripciones de escenarios: Se realiza un seguimiento continuo de la efectividad de los escenarios a través de las métricas recopiladas, refinando y actualizando los escenarios en función de las necesidades cambiantes y el análisis del rendimiento detallado por las actualizaciones del PGD hechas por el investigador principal.

## 2.6. METODOLOGÍA DETALLADA PARA LA DEFINICIÓN Y LA VALIDACIÓN DE ESCENARIOS DE INVESTIGACIÓN

La presente metodología se centra en la definición, la formalización, la automatización y el seguimiento continuo de escenarios de investigación claros y medibles, cruciales para el avance y la validación de proyectos científicos. Cada etapa se describe en detalle a continuación.

### 2.6.1. *Descubrimiento y redacción de escenarios de investigación claros*

El punto de partida radica en la identificación exhaustiva de los comportamientos y las interacciones clave que los científicos llevan a cabo dentro del contexto de su investigación. Estos comportamientos, a menudo implícitos o tácitos, deben ser desglosados y traducidos a escenarios concretos y detallados. Para asegurar la claridad y la eliminación de cualquier ambigüedad, se adopta la sintaxis Gherkin, un lenguaje de especificación basado en el comportamiento (Behavior-Driven Development - BDD).

La sintaxis Gherkin sigue una estructura lógica y fácil de comprender, compuesta por las siguientes palabras clave:

- *Característica (Feature)*: Describe la funcionalidad de alto nivel que se está especificando.
- *Escenario (Scenario)*: Ilustra un ejemplo específico de cómo se utiliza la funcionalidad.
- *Dado (Given)*: Establece el contexto inicial o las precondiciones antes de la acción.
- *Cuando (When)*: Describe la acción o el evento principal que ocurre.
- *Entonces (Then)*: Define el resultado esperado o las postcondiciones después de la acción.
- *Y (And), Pero (But)*: Se utilizan para añadir condiciones o resultados adicionales dentro de un mismo paso (Dado, Cuando, Entonces).

La aplicación rigurosa de esta sintaxis garantiza que cada escenario sea inequívoco y sirva como una base sólida para las etapas posteriores. Al describir los escenarios con un lenguaje natural, pero estructurado, se facilita la comunicación y la comprensión entre todos los miembros del equipo de investigación, independientemente de su nivel de especialización técnica.

### 2.6.2. Ejemplo de un escenario redactado con sintaxis Gherkin

- + *Característica*: Validación de la hipótesis sobre la influencia de la variable A en la variable B.
- + *Escenario*: Incremento significativo de la variable A resulta en un aumento medible de la variable B.
- + *Dado que* la variable A se establece en un valor inicial de 'X' y la variable B se mide inicialmente como 'Y'.
- + *Cuando* la variable A se incrementa en un 'Z' por ciento;
- + *Entonces* la variable B debería aumentar en al menos un 'W' por ciento.

## 2.7. FORMULACIÓN DE MÉTRICAS A PARTIR DE ESTOS ESCENARIOS

Una vez que los escenarios de investigación han sido definidos con claridad, el siguiente paso crucial consiste en formular métricas objetivas y medibles que permitan evaluar el éxito de cada escenario.

Estas métricas deben adherirse a los principios SMART:

- *Específicas (Specific)*: La métrica debe estar claramente definida y enfocada en un aspecto particular del escenario.
- *Medibles (Measurable)*: Debe ser posible cuantificar la métrica utilizando datos concretos.
- *Alcanzables (Achievable)*: La meta definida por la métrica debe ser realista dentro del contexto de la investigación.
- *Relevantes (Relevant)*: La métrica debe estar directamente relacionada con el objetivo del escenario y la pregunta de investigación general.

- *Temporales (Time-bound)*: Aunque no siempre aplicable directamente a la definición de la métrica en sí, el marco de evaluación de la métrica debe estar definido en el tiempo.

Para cada escenario identificado, se deben definir uno o varios resultados medibles que permitan determinar si el comportamiento esperado se ha cumplido. Estas métricas proporcionan una base objetiva para evaluar el progreso y el éxito de la investigación en relación con los objetivos establecidos.

### 2.7.1. *Ejemplo de métricas para el escenario anterior*

- M1: Porcentaje de aumento de B después del incremento de A.
- M2: Nivel de significancia estadística (p-valor) de la correlación entre el incremento de A y el aumento de B.

### 2.7.2. *Automatización de pruebas para escenarios*

La formalización de los escenarios mediante la sintaxis Gherkin facilita enormemente su transformación en pruebas automatizadas. Estas pruebas se ejecutan de forma continua, lo que permite una validación temprana y frecuente del sistema o del modelo de investigación en relación con los objetivos e hitos definidos.

La automatización de pruebas ofrece múltiples beneficios:

- *Validación continua*: Permite detectar problemas o desviaciones de los resultados esperados de manera temprana en el ciclo de investigación.
- *Eficiencia*: Reduce la necesidad de pruebas manuales repetitivas, liberando tiempo y recursos para otras actividades de investigación.
- *Consistencia*: Garantiza que las pruebas se ejecuten siempre de la misma manera, eliminando la variabilidad asociada a las pruebas manuales.
- *Cobertura*: Permite validar un amplio rango de escenarios de manera exhaustiva.
- *Documentación viva*: Las pruebas automatizadas sirven como documentación ejecutable de los comportamientos esperados del sistema o modelo.

La implementación de estas pruebas automatizadas requiere el uso de herramientas y marcos de trabajo específicos que puedan interpretar la sintaxis Gherkin y ejecutar las acciones y las verificaciones correspondientes.

### 2.7.3. *Supervisión y repetición de las descripciones de escenarios*

La etapa final y continua de esta metodología implica la supervisión constante de la efectividad de los escenarios definidos. Esto se realiza mediante el análisis de las métricas recopiladas durante la ejecución de las pruebas automatizadas y otras actividades de investigación. El

investigador principal, a través de las actualizaciones del Plan de Gestión de Datos (PGD), juega un papel crucial en este proceso. El PGD proporciona un marco para la recopilación, el almacenamiento, la gestión y el análisis de los datos, lo que a su vez informa sobre la validez y la relevancia de los escenarios definidos.

En función del análisis del rendimiento detallado por las actualizaciones del PGD, así como de las necesidades cambiantes del proyecto de investigación, los escenarios deben ser revisados, refinados y actualizados de manera iterativa. Este proceso de retroalimentación continua asegura que los escenarios sigan siendo relevantes, precisos y efectivos para guiar la investigación hacia sus objetivos.

La repetición de las descripciones de los escenarios puede implicar:

- *Aclarar o refinar el lenguaje*: Si se identifican ambigüedades o falta de claridad en la redacción original.
- *Ajustar las métricas*: Si las métricas definidas no son adecuadas para medir el éxito del escenario o si se identifican métricas más relevantes.
- *Añadir o eliminar escenarios*: En función de los nuevos descubrimientos o cambios en el enfoque de la investigación.
- *Modificar las precondiciones o los resultados esperados*: Si la comprensión del sistema o del fenómeno investigado evoluciona.

Este ciclo continuo de definición, medición, automatización y revisión de escenarios es fundamental para garantizar la rigurosidad, la validez y el progreso efectivo de la investigación científica. Un ejemplo concreto para la evaluación del subprincipio «(F1) Capacidad de búsqueda», se puede leer en sintaxis Gherkin como sigue:

- + *Dado*: «un conjunto de datos» con «metadatos»,
- + *Cuando*: «una máquina consulta los metadatos» para «recopilar (para reutilizar)» el «conjunto de datos»,
- + *Entonces*: los «metadatos requeridos» deben encontrarse y ser resolubles, por ejemplo, DOI (identificador de objeto digital).

En la siguiente sección, se llevará a cabo un análisis de los resultados que ya se han observado tras la implementación del BDD como metodología para redactar escenarios (Cucumber.io, s. f.-a) mediante sintaxis Gherkin (Cucumber.io, s. f.-b), los requerimientos por evaluar en los resultados de investigación, independientemente de su naturaleza o disciplina, sino enfocados en su tipo de activo y cómo este se adhieren a los principios FAIR (Wilkinson *et al.*, 2016).

## 2.8. ANÁLISIS DE RESULTADOS CON EL BDD Y PRINCIPIOS FAIR

Esta implementación se ha centrado específicamente en la formulación de los requerimientos de evaluación de resultados de investigación a través de la sintaxis Gherkin. Cabe destacar que este enfoque metodológico se aplica de manera transversal, trascendiendo las particularidades inherentes a la naturaleza o disciplina de cada investigación. En lugar de

segmentar el análisis por campos específicos del conocimiento, la atención se centra en la tipología del activo resultante de la investigación y la manera en que estos activos se alinean y cumplen con los pilares fundamentales de los principios FAIR (Wilkinson *et al.*, 2016).

Este análisis detallado permitirá obtener una comprensión profunda del impacto de la metodología BDD (Cucumber.io, s. f.-a) en la claridad, precisión y verificabilidad de los requerimientos de evaluación. Asimismo, se explorará cómo la adopción de un lenguaje común y estructurado como Gherkin (Cucumber.io, s. f.-b) facilita la comunicación y la colaboración entre los diferentes actores involucrados en el proceso de investigación y evaluación, incluyendo investigadores, desarrolladores y otros *stakeholders* (Agile Alliance, s. f.).

La evaluación de la adherencia a los principios FAIR se realizará para cada tipo de activo identificado, analizando en qué medida los resultados de investigación, independientemente de su formato (datos, software, publicaciones, etc.), cumplen con los criterios específicos de cada principio.

- En relación con la *Findability*, se examinará la calidad de los metadatos asociados y la facilidad con la que los activos pueden ser descubiertos mediante búsquedas.
- En cuanto a la *Accessibility*, se analizarán los protocolos de acceso, las licencias y las posibles restricciones existentes.
- Respecto a la *Interoperability*, se evaluará la utilización de formatos y estándares abiertos que permitan la combinación y el intercambio de datos con otros activos.
- Finalmente, en lo referente a la *Reusability*, se considerarán las licencias de uso, la documentación detallada y la proveniencia de los datos que faciliten su posterior utilización en diferentes contextos.

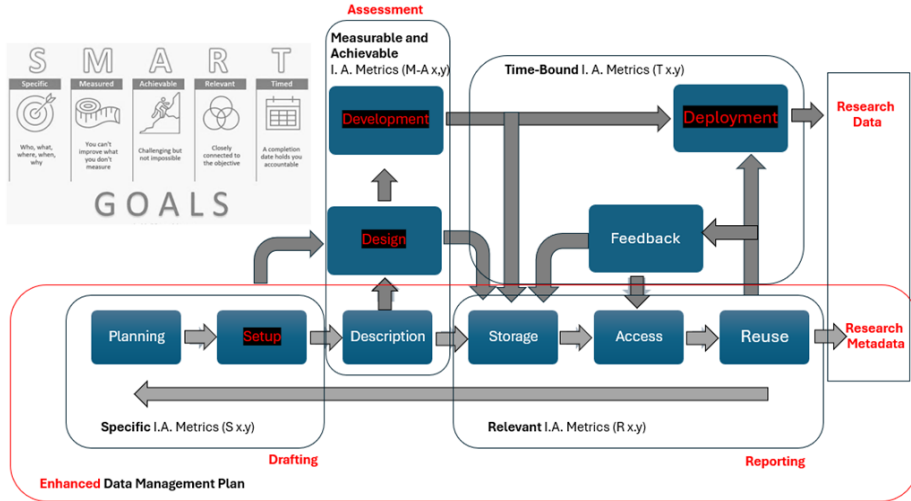
A través de este análisis estructurado se busca no solo identificar el grado de cumplimiento de los principios FAIR (Wilkinson *et al.*, 2016) de los resultados de investigación redactados con la metodología BDD (Cucumber.io, s. f.-a), sino también destacar las ventajas de este enfoque para mejorar la calidad, la transparencia y el impacto de la investigación en general.

Los hallazgos iniciales de la siguiente sección proporcionarán información valiosa para la optimización de los procesos de investigación y la promoción de prácticas más abiertas y colaborativas en el ámbito científico y académico.

### 3. RESULTADOS Y DISCUSIÓN

Si bien este artículo se centra en la presentación del marco conceptual y la metodología, es importante destacar los resultados esperados y los beneficios potenciales de su implementación.

Figura 2. Métricas SMART adaptadas para RDLC para proveer metadatos adicionales para FAIRness (Grasso Toro, 2024b)



Se presentarán los casos de uso iniciales y se explorarán algunos detalles de los cambios y mejoras tangibles que se han identificado en la gestión y la difusión de los resultados de investigación como consecuencia directa de la aplicación de esta estrategia.

### 3.1. CASOS DE USO INICIALES

El marco propuesto tiene un amplio potencial de aplicación en diversos dominios de investigación. Algunos casos de uso en desarrollo incluyen:

- Ensayos clínicos: Se está explorando la aplicación del marco para el seguimiento de datos sensibles en ensayos clínicos, asegurando el cumplimiento del Reglamento General de Protección de Datos (RGPD) y la correcta compartición de métodos para la replicabilidad.
- Ciencias ambientales: Se está investigando la aplicación del marco para la monitorización de cambios ecológicos a largo plazo y la facilitación del intercambio de datos, modelos y ontologías.
- Humanidades digitales: Se está evaluando la aplicación del marco para la gestión de conjuntos de datos complejos de materiales textuales y multimedia para su análisis y preservación.

El marco propuesto demuestra una versatilidad considerable al ser aplicado y evaluado en una variedad de campos con necesidades específicas de gestión y compartición de datos sensibles y complejos.

### 3.1.1. *Ensayos clínicos*

La aplicación del marco en el ámbito de los ensayos clínicos se centra en abordar los desafíos inherentes al manejo de datos sensibles de pacientes. Esto incluye la recopilación, el procesamiento, el almacenamiento y la compartición de información clínica y resultados de investigación. Un aspecto crucial es garantizar el estricto cumplimiento del Reglamento General de Protección de Datos (RGPD) y otras normativas de privacidad relevantes.

El marco busca facilitar la implementación de mecanismos robustos de anonimización, seudonimización y control de acceso para proteger la confidencialidad de los participantes. Además, se explora cómo el marco puede estructurar y documentar los métodos utilizados en los ensayos clínicos, permitiendo una compartición transparente y facilitando la replicabilidad de los resultados por parte de otros investigadores. Esto promueve la transparencia y la colaboración en la investigación médica, acelerando el avance del conocimiento y la identificación de tratamientos efectivos. La capacidad del marco para gestionar la trazabilidad de los datos y los metadatos asociados también es fundamental para asegurar la integridad y la auditabilidad de los ensayos clínicos.

### 3.1.2. *Ciencias ambientales*

En el campo de las ciencias ambientales, el marco se está investigando como una herramienta para mejorar la monitorización de cambios ecológicos a largo plazo. Esto implica la gestión de grandes volúmenes de datos heterogéneos provenientes de diversas fuentes, como sensores remotos, estaciones de monitoreo terrestre y observaciones de campo. El marco busca facilitar la integración y el análisis de estos datos para identificar tendencias, patrones y anomalías en los ecosistemas. Un aspecto fundamental es la promoción del intercambio de datos entre diferentes instituciones y grupos de investigación, superando las barreras técnicas y semánticas que a menudo dificultan la colaboración. Para ello, se está explorando cómo el marco puede soportar el uso de modelos de datos y ontologías compartidas, asegurando la interoperabilidad y la comprensión común de la información. La capacidad del marco para gestionar la procedencia de los datos y los metadatos relevantes es esencial para garantizar la calidad y la fiabilidad de los análisis ecológicos. Esto permite una mejor comprensión de los impactos ambientales y la formulación de estrategias de conservación más efectivas.

### 3.1.3. *Humanidades digitales*

El ámbito de las humanidades digitales se caracteriza por el creciente uso de conjuntos de datos complejos (i. e., materiales textuales, imágenes, audio y video). El desarrollo del marco se evalúa como una solución para la gestión, el análisis y la preservación de estos recursos digitales. Los desafíos incluyen la diversidad de formatos de datos, la necesidad de herramientas de análisis especializadas y la importancia de la preservación a largo plazo para futuras generaciones de investigadores. El marco podría facilitar la organización y la catalogación de

estos conjuntos de datos, permitiendo un acceso más eficiente y la aplicación de técnicas de análisis avanzado, como el procesamiento del lenguaje natural y la visión por computador. Además, se está considerando su potencial para apoyar la creación de repositorios digitales interoperables, facilitando el descubrimiento y la reutilización de materiales de investigación. La capacidad del marco para gestionar metadatos ricos y la información sobre los derechos de autor y las condiciones de uso es crucial para garantizar el acceso ético y sostenible a estos valiosos recursos culturales e históricos.

### 3.2. OBSERVACIONES INICIALES A NIVEL DE PROYECTO

La aplicación del desarrollo dirigido por el comportamiento (BDD) para formalizar la evaluación de la FAIRness ya está mostrando resultados prometedores. Al definir escenarios de uso específicos y criterios de evaluación claros, el BDD no solamente se le permite al investigador principal una evaluación más precisa y objetiva de la FAIRness (capacidad de ser Encontrables, Accesibles, Interoperables y Reutilizables) de los resultados, sino también de otros activos de investigación cruciales, como son protocolos, código computacional e incluso ontologías reutilizadas de proyectos anteriores.

El uso del BDD para evaluar la FAIRness de los datos de investigación y otros activos clave está mostrando resultados prometedores. Al definir escenarios de uso y criterios claros, el BDD permite una evaluación más precisa, facilita la identificación de áreas de mejora y proporciona un elemento reutilizable para la implementación de soluciones que optimizan la FAIRness. Al alinear las expectativas con las capacidades, el BDD fomenta la colaboración entre los equipos y acelera el progreso científico.

### 3.3. MEJORA DE LA COMUNICACIÓN MÁQUINA A MÁQUINA

El marco propuesto mejora la gestión de datos de investigación, facilitando la comunicación entre máquinas y permitiendo la automatización de tareas, evaluación de datos, generación de informes y seguimiento del progreso.

Uno de los principales beneficios de este marco es la optimización de la comunicación entre máquinas. Al integrar las métricas SMART y los principios FAIR en los PGD, se crea un lenguaje común que facilita el intercambio de información entre diferentes sistemas y plataformas. Esto ha permitido la automatización de tareas a nivel de grupo de investigación, la evaluación trazable de la FAIRness de los datos a nivel de trabajo del gestor de datos, la generación de informes a nivel de oficina de gestión de fondos de investigación y el seguimiento del progreso de la investigación.

La implementación de esta infraestructura ha facilitado la automatización de diversas tareas operativas dentro del grupo de investigación, optimizando la eficiencia y reduciendo la carga de trabajo manual. Esto incluye, entre otros aspectos, la gestión de metadatos, la organización de archivos y la ejecución de flujos de trabajo predefinidos.

Adicionalmente, se ha establecido un sistema de evaluación trazable para garantizar la conformidad de los datos de investigación con los principios FAIR (Findable, Accessible, Interoperable, Reusable) a lo largo de todo el ciclo de vida de los datos, con una supervisión detallada a nivel del gestor de datos responsable. Esta trazabilidad permite identificar y corregir posibles desviaciones de los estándares de calidad, asegurando la integridad y el potencial de reutilización de los datos generados.

En cuanto a la gestión administrativa, la plataforma posibilita la generación automática de informes personalizados, adaptados a las necesidades de la oficina de gestión de fondos de investigación. Estos informes proporcionan una visión clara y actualizada del estado de los proyectos, facilitando la toma de decisiones y la justificación de la inversión realizada.

Finalmente, la herramienta implementada ofrece funcionalidades de seguimiento del progreso de la investigación en tiempo real, permitiendo a los investigadores y a los responsables del grupo visualizar los avances, identificar posibles obstáculos y ajustar las estrategias según sea necesario. Este seguimiento continuo contribuye a una gestión más eficaz de los proyectos y a la optimización de los resultados de la investigación.

### 3.4. OPTIMIZACIÓN DEL TIEMPO Y AUMENTO DE LA TRANSPARENCIA

La transparencia y la trazabilidad que promueve este marco se traducen en una mayor eficiencia en la gestión de la investigación. Al eliminar ambigüedades y establecer criterios claros y medibles, se reducen los tiempos de revisión y aprobación de los PGD. Además, la automatización de ciertas tareas relativas al mantenimiento de datos y la disponibilidad de información en tiempo real permiten a los investigadores y gestores de datos tomar decisiones más informadas y oportunas.

### 3.5. RESULTADOS ESPECÍFICOS Y DESARROLLO CONTINUO

Es importante mencionar que, aunque los PGD mejorados con este marco están todavía en desarrollo, se espera que su implementación tenga un impacto significativo en la calidad y la eficiencia de la investigación. Se prevé que los PGD mejorados faciliten la evaluación de la FAIRness de los datos, promuevan la reutilización de los datos y optimicen la comunicación entre investigadores, instituciones académicas y entidades de financiación, como se ilustra en la Figura 3.

## 4. CONCLUSIONES

El marco propuesto ofrece una solución integral para mejorar el manejo de los PGD y promover la ciencia abierta. Al integrar las métricas SMART, los principios FAIR, el RDLC y el BDD, este marco facilita la creación de PGD más robustos, transparentes y eficientes.

Si bien los resultados específicos de su implementación aún están por verse, se espera que este marco tenga un impacto positivo en la calidad, la eficiencia y la transparencia de la investigación.

#### 4.1. RESUMEN DEL MARCO INTEGRAL PARA EL MANEJO DE PLANES DE GESTIÓN DE DATOS

El marco conceptual propuesto representa una estrategia comprehensiva y multifacética destinada a optimizar sustancialmente el manejo de los planes de gestión de datos (PGD) y, simultáneamente, impulsar la adopción y los principios de la ciencia abierta dentro de la comunidad investigadora. Su principal fortaleza reside en la integración sinérgica de diversas metodologías y principios fundamentales, incluyendo las métricas SMART (Específicas, Medibles, Alcanzables, Relevantes y con plazos definidos de tiempo), el ciclo de vida de los datos de investigación (RDLC), los principios FAIR (Encontrables, Accesibles, Interoperables y Reutilizables) y la descripción de datos básicos (BDD).

La incorporación de las métricas SMART dentro de este marco tiene como objetivo asegurar que los objetivos establecidos en los PGD sean claramente definidos, susceptibles de seguimiento y evaluación a lo largo del ciclo de vida del proyecto de investigación. Esto permite a los investigadores y a las instituciones monitorizar el progreso en la gestión de datos y realizar ajustes informados según sea necesario.

El ciclo de vida de los datos de investigación (RDLC) proporciona una estructura lógica para la gestión de datos, desde su planificación y recopilación hasta su procesamiento, análisis, preservación y reutilización. Al integrar el RDLC en el marco propuesto, se asegura que los PGD aborden todas las etapas cruciales del ciclo de vida de los datos, promoviendo una gestión proactiva y reflexiva de los recursos de datos.

La adhesión a los principios FAIR es un componente central de este marco, reconociendo la creciente importancia de garantizar que los datos de investigación sean fácilmente descubiertos por otros investigadores, accesibles bajo condiciones claras y comprensibles, interoperables con diferentes conjuntos de datos y sistemas y susceptibles de ser reutilizados para futuras investigaciones. La implementación de los principios FAIR no solo aumenta la visibilidad y el impacto de los datos de investigación, sino que también fomenta la colaboración y la reproducibilidad de los resultados científicos.

La inclusión de la descripción basada en comportamiento del sistema de manejo de datos (mediante el BDD) dentro del marco subraya la necesidad de documentar adecuadamente los datos de investigación, proporcionando información esencial sobre su contexto, metodología de recopilación, estructura y posibles limitaciones, siguiendo los escenarios descritos con sintaxis Gherkin. Una documentación exhaustiva, facilitada por el BDD, mejora significativamente la comprensibilidad y la reutilización de los datos por parte de la comunidad científica.

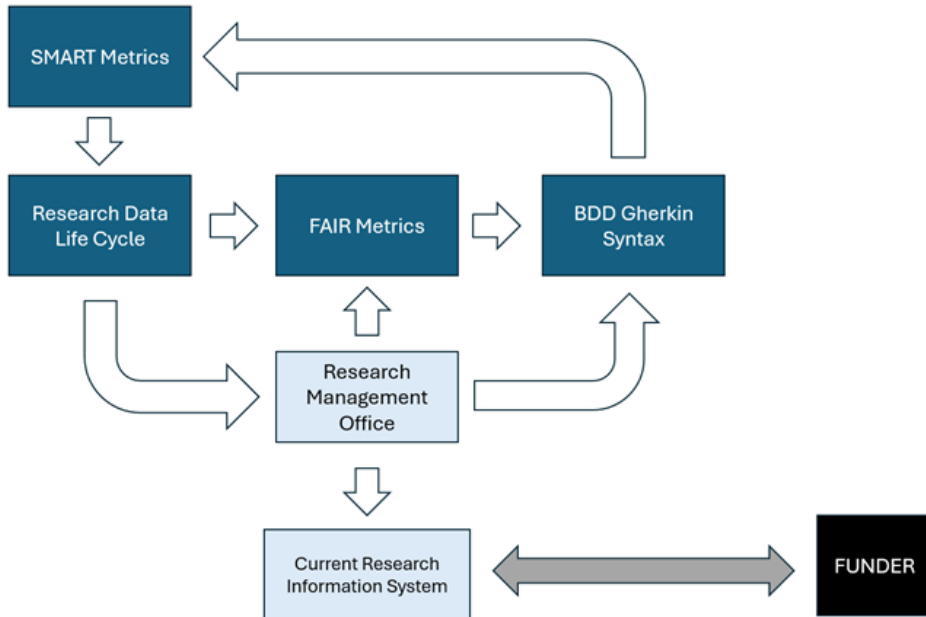
En conjunto, la integración de las métricas SMART, los principios FAIR, el RDLC y el BDD en este marco propuesto crea un ecosistema favorable para la creación de PGD que son

inherentemente más robustos en su planificación, transparentes en su ejecución y eficientes en la gestión de los recursos de datos.

Si bien es crucial reconocer que los resultados concretos y cuantificables de la adopción de este marco se manifestarán plenamente a través de su implementación práctica y la evaluación de su impacto a largo plazo, las expectativas iniciales sugieren un efecto considerablemente positivo en aspectos fundamentales de la investigación científica.

Se anticipa una mejora notable en la calidad de los datos gestionados, un aumento en la eficiencia de los procesos de investigación gracias a una mejor planificación y organización de los datos y un incremento significativo en la transparencia de la investigación, facilitando la verificación de los resultados y la construcción de conocimiento sobre bases más sólidas. En última instancia, este marco aspira a contribuir de manera sustancial a la evolución hacia una ciencia más abierta, colaborativa y de mayor impacto.

Figura 3. Marco de mejora para la comunicación entre investigadores, instituciones y entidades de financiación (Grasso Toro, 2024b)



#### 4.2. EL ROL DEL GESTOR DE DATOS EN LA IMPLEMENTACIÓN DE ESTE MARCO

La implementación exitosa del marco propuesto depende en gran medida de la participación activa de los gestores de datos (*Data Stewards* en inglés).

Estos profesionales desempeñan un papel crucial en la recopilación de requisitos, la colaboración entre los diferentes actores y la validación de las métricas SMART de los activos intelectuales por investigar. El marco propuesto tiene el potencial de ser aplicado en una amplia gama de dominios de investigación, incluyendo ensayos clínicos, ciencias ambientales y humanidades digitales, como se mencionó anteriormente.

La capacidad para fomentar una mayor transparencia, reproducibilidad y colaboración entre investigadores, instituciones académicas y entidades de financiación posiciona este marco como una herramienta de gran valor para impulsar el avance de la reutilización del conocimiento científico (Grasso Toro, 2024b).

Además, puede facilitar la gestión y el intercambio eficiente de datos de investigación, promoviendo la integridad de los datos y reduciendo la duplicación de esfuerzos. Tiene el potencial de democratizar el acceso a la información científica y acelerar el descubrimiento y la divulgación, convirtiéndose en un elemento clave para el futuro de la investigación y a su vez consolidar nuevas iniciativas como la ciencia ciudadana. La implementación exitosa de este marco requerirá la participación proactiva de investigadores, gestores de datos y entidades de financiación.

## REFERENCIAS BIBLIOGRÁFICAS

- Agile Alliance. (s. f.). *Three amigos*. <https://www.agilealliance.org/glossary/three-amigos/>
- Burnette, M. H., Williams, S. C. e Imker, H. J. (2016). From plan to action: Successful data management plan implementation in a multidisciplinary project. *Journal of eScience Librarianship*, 5(1), e1101. <https://doi.org/10.7191/jeslib.2016.1101>
- Corporate Finance Institute. (s. f.). *SMART goal - Definition, guide, and importance of goal setting*. <https://corporatefinanceinstitute.com/resources/management/smart-goal/>
- Cucumber.io. (s. f.-a). *Who writes the scenarios?* <https://cucumber.io/blog/bdd/who-writes-the-cucumber-scenarios/>
- Cucumber.io. (s. f.-b). *Gherkin reference*. <https://cucumber.io/docs/gherkin/reference/>
- Dataversity. (s. f.). *A brief history of data management*. <https://www.dataversity.net/brief-history-data-management/>
- Grasso Toro, F. (2024). DMPs as Management Tools for Intellectual Assets by SMART-metrics. *International Journal of Digital Curation*, 18(1). <https://doi.org/10.2218/ijdc.v18i1.919>
- Grasso Toro, F. (2024, May 13-17). *Assisting FAIRness lifecycle: SMART-metrics for FAIR research data lifecycle management* [Presentación en Power Point]. EnhanceR Symposium 2024, Zurich, Switzerland. <https://symposium2024.enhancer.ch/materials/>
- Schmitt, R. H., Bodenbenner, M., Hamann, T., Sanders, M. P., Moser, M. y Abdelrazeq, A. (s. f.). *Leveraging measurement data quality by adoption of the FAIR guiding principles*. [Preprint Manuscript].
- Wilkinson, M. D. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>



## AS PRODUÇÕES CIENTÍFICAS BRASILEIRAS SOBRE A INFORMAÇÃO PARA A SUSTENTABILIDADES NA CIÊNCIA A LUZ DO PROJETO «ESTRUTURAÇÃO DA INFORMAÇÃO SOBRE A PESQUISA BRASILEIRA»


### *BRAZILIAN SCIENTIFIC PRODUCTIONS ON INFORMATION FOR SUSTAINABILITY IN INFORMATION SCIENCES IN LIGHT OF THE PROJECT «STRUCTURING INFORMATION ON BRAZILIAN RESEARCH»*

Ana Suely PINHO LOPES

*Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Brasil*

*Universidade de Brasília. Brasil*

[pinholopes.anasuely@gmail.com](mailto:pinholopes.anasuely@gmail.com)

 <https://orcid.org/0000-0001-5047-7315>

**RESUMO:** As produções científicas brasileiras no contexto da informação para a sustentabilidade têm sido publicadas em diversas áreas da produção do conhecimento científico nas esferas pública, social, econômica e científica. A sustentabilidade é percebida como solução para manter o equilíbrio ambiental por meio do reuso dos recursos disponíveis. Desde então, o desafio das organizações passou a ser o de buscar boas práticas para atingir seus objetivos. Com a Ciência da Informação ocorre o mesmo, a informação para a sustentabilidade é vista como um recurso fundamental no uso e acesso à informação. Para tanto, necessita ser disponibilizada por meio de uma gestão eficiente para atender as demandas da sociedade em geral. O presente artigo tem como objetivo identificar as produções intelectuais e científicas brasileiras da Ciência da Informação sobre a informação para a sustentabilidade. Foi feita uma revisão bibliográfica tendo por base produções científicas publicadas em forma de artigos, dissertações e teses. As bases de dados BDTD, BDLIS e UFSC foram selecionadas para estabelecer o corpus da revisão de literatura. Para tanto, utilizado como critério 67 publicações científicas brasileiras a incluir pesquisas que abordam sobre a informação para a sustentabilidade na Ciência da Informação. Esse processo resultou em 17 produções lidas e interpretadas desde a análise de conteúdo. Identificou-se que a produção científica brasileira na área da Ciência da Informação sobre a informação para a sustentabilidade é um campo ainda pouco explorado pelos pesquisadores científicos e concentra-se nas áreas de administração, saúde, contabilidade. Nessa área, portanto, são recorrentes aos conceitos, ou seja, a tentar compreender a importância do desenvolvimento

Ana Suely Pinho Lopes

*As produções científicas brasileiras sobre a informação para a sustentabilidades na ciência a luz do projeto “estruturação da informação sobre a pesquisa brasileira”*

de tecnologias sustentáveis. Foi possível inferir que a produção científica brasileira na área da Ciência da Informação sobre a informação para a sustentabilidade está sendo explorada, porém, em nível restrito, mas que já suscita incentivo para a elaboração de saberes e fazeres para o seu crescimento.

*PALAVRAS-CHAVE:* produção científica; ciência da informação; informação; sustentabilidade.

**ABSTRACT:** Brazilian scientific productions in the context of information for sustainability have been published in several areas of scientific knowledge production in the public, social, economic and scientific spheres. Sustainability is perceived as a solution to maintain environmental balance through the reuse of available resources. Since then, the challenge for organizations has become to seek good practices to achieve their goals. The same occurs with Information Science, information for sustainability is seen as a fundamental resource in the use and access to information. Therefore, it needs to be made available through efficient management to meet the demands of society in general. This article aims to identify Brazilian intellectual and scientific productions in Information Science on information for sustainability. A bibliographic review was carried out based on scientific productions published in the form of articles, dissertations and theses. The BDTD, BDLIS and UFSC databases were selected to establish the corpus of the literature review. To this end, 67 Brazilian scientific publications were used as criteria to include research that addresses information for sustainability in Information Science. This process resulted in 17 productions that were read and interpreted from the content analysis. It was identified that Brazilian scientific production in the area of Information Science on information for sustainability is a field that is still little explored by scientific researchers and is concentrated in the areas of administration, health, and accounting. In this area, therefore, concepts are recurrent, that is, trying to understand the importance of developing sustainable technologies. It was possible to infer that Brazilian scientific production in the area of Information Science on information for sustainability is being explored, however, at a restricted level, but that it already encourages the development of knowledge and actions for its growth.

**KEYWORDS:** scientific production; information science; information; sustainability.

**RESUMEN:** Las producciones científicas brasileñas en el contexto de la información para la sostenibilidad han sido publicadas en diversas áreas de la producción de conocimiento científico en las esferas pública, social, económica y científica. La sostenibilidad se percibe como una solución para mantener el equilibrio ambiental mediante la reutilización de los recursos disponibles. Desde entonces, el reto de las organizaciones ha sido buscar buenas prácticas para alcanzar sus objetivos. Lo mismo ocurre con la ciencia de la información, la información para la sostenibilidad es vista como un recurso fundamental en el uso y acceso a la información. Para ello es necesario ponerlo a disposición mediante una gestión eficiente para satisfacer las demandas de la sociedad en general. Este artículo tiene como objetivo identificar las producciones intelectuales y científicas brasileñas en ciencia de la información sobre información para la sostenibilidad. Se realizó una revisión bibliográfica con base en la producción científica publicada en forma de artículos, disertaciones y tesis. Se seleccionaron las bases de datos BDTD, BDLIS y UFSC para establecer el corpus de la revisión de la literatura. Para ello, se utilizaron

67 publicaciones científicas brasileñas como criterio para incluir investigaciones que aborden la información para la sostenibilidad en la ciencia de la información. Este proceso dio como resultado 17 producciones leídas e interpretadas a partir del análisis de contenido. Se identificó que la producción científica brasileña en el área de ciencias de la información sobre información para la sostenibilidad es un campo aún poco explorado por los investigadores científicos y está concentrada en las áreas de administración, salud y contabilidad. En este ámbito, por tanto, los conceptos son recurrentes, es decir, tratar de comprender la importancia de desarrollar tecnologías sostenibles. Se pudo inferir que la producción científica brasileña en el área de ciencias de la información sobre información para la sostenibilidad viene siendo explorada, sin embargo, a un nivel restringido, pero que ya proporciona incentivo a la elaboración de conocimientos y acciones para su crecimiento.

PALABRAS CLAVE: producción científica; ciencias de la información; información; sostenibilidad.

## 1. INTRODUÇÃO

Considere-se que um dos maiores desafios para uma organização contemporânea é gerenciar a informação produzida no ambiente interno e externo das organizações. Nesse cenário, a Ciência da Informação (CI) enquanto uma área que estuda, gerencia e organiza meios de tratamento, disseminação, recuperação e uso da informação, tem como desafio o de inserir em seu programa o estudo e desenvolvimento sobre a informação para a sustentabilidade, dada a necessidade de suprir as necessidades informacionais de uma sociedade sustentável.

Sabe-se que as esferas da educação ambiental juntamente com a temática da sustentabilidade e da responsabilidade socioambiental constituem-se um cenário para implementar um repositório destinado ao saberes e fazeres voltado para a educação e política ambiental que vai ao encontro da gestão da informação, do conhecimento por meio de projetos, com compromisso, integridade e perseverança – fatores fundamentais para o desenvolvimento de pesquisas e inovação na CI voltados à formação de um cidadão cada vez mais consciente e de atitudes e práticas inovadoras diante de suas responsabilidades para com o meio ambiente, desde as políticas públicas que têm como um dos principais propósitos a sustentabilidade.

Nessa perspectiva, considera-se importante visualizar a informação para a sustentabilidade como um elo de intercessão entre as dimensões ambiental, econômica e social da sustentabilidade. Ressalta-se a importância de incluir na área da CI a informação para a sustentabilidade, como um veículo de sensibilização, conscientização e mobilização da sociedade como um todo, seja ela científica ou civil, a tratar das práticas, dos objetivos e metas de desenvolvimento capaz de suprir as necessidades da sociedade atual, sem comprometer a capacidade de atender as necessidades das futuras gerações.

Para Santos e Cândido (2022) a possibilidade de um modelo de desenvolvimento sustentável a se tratar da esfera internacional, requer, portanto, um sistema de informações robusto tanto em qualidade quanto em quantidade, cuja idealização e formação requer esforços dos

profissionais da área e paralelamente comprometimento dos diversos atores envolvidos com a gestão do projeto da sociedade.

Pressupõe-se nesse cenário, que a informação para a sustentabilidade tem como objetivo facilitar tanto o acesso quanto a utilização da informação para alcançar o desenvolvimento sustentável, isso vale tanto para o contexto atual visando a colaboração para com a preservação dos recursos naturais, da vida humana, do combate às desigualdades sociais, assim como para assegurar prosperidade e garantir o acesso à informação às gerações futuras (Lopes, 2017).

Nesse horizonte, o incentivo para o aumento das produções científicas brasileiras sobre a informação para a sustentabilidade, nomeadamente na CI, contribuirá mais e mais para com os estudos dessa questão, proporcionando induzir a produção científica nessa área a tornar-se um contributo para o alcance de objetivos na prática da informação para a sustentabilidade nas diversas esferas sociais. Sendo assim, o objetivo desse artigo consiste em identificar a existência das produções científicas brasileiras publicadas sobre a informação para a sustentabilidade na área da CI.

### 1.1. A SUSTENTABILIDADE NA CIÊNCIA DA INFORMAÇÃO

Com o célere avanço tecnológico sob o domínio de todas as áreas, foi que no começo desse século, o avanço da tecnologia de informação adentrava com foco na sustentabilidade com a pauta de discussões na Ciência da Informação (CI), fortalecendo a proposta do professor e pesquisador Dr. Jan Nolin, da Escola Sueca de Biblioteconomia e Ciência da Informação de Borås (Suécia), que em 2010 sugeriu o uso de tecnologias de informação direcionada à sustentabilidade e, sincronicamente, para contribuir com objetivos globais.

Na perspectiva de Nolin (2010) as vertentes que introduzem a argumentação da sustentabilidade na CI é formada pelo seguinte tripé: Desenvolvimento de tecnologia de informação sustentável: aspirando à redução da divisão digital (desenvolvimento social) por meio da transferência de tecnologia e desenvolvimento econômico, nivelando o acesso à informação para todas as pessoas; Desenvolvimento da conservação sustentável da informação: propondo reunir e estruturar informações sobre diversos tipos de indicadores sustentáveis de modo que torne fácil recuperá-las e visualizá-las de diferentes maneiras com qualidade e garantia da informação e desenvolvimento do compartilhamento de informações sustentáveis: ao perceber que além de permitir o acesso e uso da informação, todas as pessoas de variadas culturas devem ser incitadas a participar efetivamente dos processos de tomada de decisão, e no; Desenvolvimento da cultura de sentido de pertencimento coletivo para a evolução e efetivação do desenvolvimento sustentável.

Nessa conjuntura, o termo *sustentabilidade* pode ser aceito como a capacidade de desenvolver e utilizar a informação de maneira sustentável, de forma que ela seja gerenciada de jeito eficiente e responsável, de modo que não venha a esgotar seus recursos ou comprometer a integridade dos sistemas que a usufruem. Para além, como instrumento utilizado para sensibilizar, conscientizar e mobilizar toda sociedade sobre a importância de apoiar objetivos universais de desenvolvimento sustentável.

Para tanto, um dos principais desafios da informação para a sustentabilidade é lidar com a quantidade crescente de dados gerados e utilizados no dia a dia. Sendo assim, é necessário adotar práticas de gestão de dados mais eficientes e sustentáveis, a citar, a compressão de arquivos, a duplicação de informações, a utilização de algoritmos de compressão de dados e o armazenamento em nuvem.

Ademais, a informação para a sustentabilidade relaciona-se a preservação da memória e do conhecimento produzido e armazenado, garantindo que informações importantes sejam asseguradas para a recuperação, acesso e uso pelas futuras gerações. Nesse sentido inclui-se desde a conservação e preservação de registros históricos dos feitos da humanidade até o desenvolvimento de sistemas mais robustos de backup e recuperação de dados a considerar a fragilidade e obsolescência dos suportes tecnológicos em constante evolução e transformação.

Para além, a informação para a sustentabilidade deve ser percebida como um aparato de apoio ao acesso à informação com desenvolvimento sustentável, e assim, sensibilizar e demonstrar à sociedade da informação, a importância da mobilização coletiva a favor de objetivos universais de desenvolvimento sustentável.

Em síntese, a informação para a sustentabilidade é uma concepção fundamental para a preservação da informação em um cenário cada vez mais interligado e tecnológico. No momento que adotarmos práticas de gestão e uso de dados mais responsáveis e eficientes, podemos assegurar a segurança dos dados, a privacidade e a disponibilidade das informações, e ao mesmo tempo contribuir para sensibilização e mobilização da sociedade da informação direcionado a um mundo mais sustentável e equilibrado, por seu turno a dar um contributo para uma melhor conservação do meio ambiente.

Nessa perspectiva, suscitam-nos os questionamentos de como fazer para desenvolver e aprimorar tecnologias sustentáveis de informação, por meio de quais ações estratégicas que venham favorecer uma obtenção informacional de forma eficaz e como a área da CI pode contribuir para a informação para a sustentabilidade. Inquietações essas que desafiaram estudiosos como Spink (1999) e Nolin (2010), em busca de desenvolver pesquisas para encontrar respostas para questões fundamentais do sistema democrático no ambiente da informação para a sustentabilidade que uma sociedade moderna e avançada tem como desafio, onde os cientistas da informação podem considerar debater a respeito de uma futura possibilidade científica de uma sociedade sustentável, no bojo de uma abordagem de necessidades estruturais para a ampliação da sustentabilidade.

Porém, no cenário das produções científicas brasileiras sobre a informação para a sustentabilidade, pesquisadores da CI vem a realizar estudos recentes para discorrer sobre essa questão, ainda que timidamente a explorar os conceitos.

Como protótipo pode-se referenciar Albagli (1995), que discorreu sobre a forma sobre a qual, o papel da informação e especificamente da «informação ambiental», no âmbito do modelo de desenvolvimento sustentável, se comportava como alternativa para a crise socioeconômica-ambiental em esfera universal.

Ao referir-se sobre a dinâmica global relacionada ao conhecimento e à sustentabilidade, Arruda (2009) enxerga nesse ambiente o acesso à informação advindo cada vez mais célere e configurando-se como a mudança do arquétipo industrial para o tecnológico.

Araújo, Inomata e Rados (2014) no que lhes respeitam, salientam que o processo de gestão da informação pode ser uma espécie de assistência para potencializar a eficiência e a respectiva eficácia da organização, avaliando o uso de informações de baixo custo, que estão disponibilizadas e que podem tornar-se estratégicas para a organização, convertendo em conhecimento para a geração de soluções inovadoras que atendam às exigências impostas alinhadas ao desenvolvimento da sustentabilidade. Por seu turno, Geraldo e Pinto (2019, 2021b) ao manifestarem-se a respeito da sustentabilidade informacional, ao validarem a importância do discurso da sustentabilidade na CI, expressando o intuito de consolidar esta temática nas futuras pesquisas científicas. Esses autores relatam ainda que o acesso à informação pode ser um meio de promover a disseminação do conhecimento sobre o que é sustentabilidade e quais os benefícios do desenvolvimento sustentável, imaginando cidadãos conscientes, críticos e atuantes em ações universais de sustentabilidade, na elaboração e monitoramento de projetos legislativos eficientes conforme os objetivos universais de bem-estar socioeconômico ambiental.

## 2. METODOLOGIA

A presente pesquisa teve como objetivo identificar as produções científicas brasileiras na Ciência da Informação (CI) sobre a informação para a sustentabilidade. Desse modo, foi realizada uma revisão bibliográfica que consiste em um estudo que tem por base trabalhos já publicados em livros, artigos, revistas, periódicos científicos, teses, dissertações e anais de eventos científicos. A pesquisa foi realizada nas bases de dados científicas Biblioteca Digital Brasileira de Teses e Dissertações<sup>1</sup> (BDTD), Biblioteca Digital do Laboratório de Informação para Sustentabilidade<sup>2</sup> (BDLIS), Universidade Federal de Santa Catarina (UFSC) e em sítios de bibliotecas científicas. Para a recuperação da temática foram utilizados os seguintes termos: Produção científica; Ciência da Informação; Informação; Sustentabilidade. Após o exame das produções recuperadas e, após análise, foram selecionadas aquelas que envolvem especificamente o assunto de interesse da pesquisa.

## 3. RESULTADOS E DISCUSSÃO

A primeira etapa referiu-se à pesquisa das produções científicas brasileiras na área de Ciência da Informação (CI) sobre a informação para a sustentabilidade de forma que fosse útil para ser depositada no repositório digital de interesse da pesquisa.

Os termos utilizados para a busca foram Ciência da Informação e sustentabilidade. A produção científica identificada foi efetuada nos sítios de bases de dados científicas, foram encontradas, na maioria, nas áreas de Administração e Contabilidade, às que se referem à CI, reportam-se a parte conceitual, sobre a conscientização da importância da sustentabilidade.

1 <https://bdt.d.ibict.br/vufind/>

2 <https://bdlis.ibict.br/>

O Quadro 1 apresenta o levantamento da produção científica brasileira apresentada por título, resumo/conclusão e fonte. Abaixo do Quadro 1 está descrita a análise onde relata-se a abordagem sobre os desafios que as organizações contemporâneas enfrentam no gerenciamento do conhecimento produzido e faz-se uma breve abordagem sobre a perspectiva da CI no contexto da sustentabilidade. Conclui-se com comentários, citação de algumas tendências dessas pesquisas e a perspectiva do impacto nas futuras pesquisas.

Quadro 1. Levantamento da produção científica brasileira

Título	Resumo/conclusão	Fonte
«Gestão do Conhecimento e Sustentabilidade no Contexto Organizacional: Uma Revisão Sistemática na Ciência da Informação» (Fonseca, Mota e Santos Junior, 2022).	«Um dos grandes desafios para uma organização contemporânea é gerenciar o conhecimento produzido no ambiente interno e externo da organização. Para isso, as organizações estão investindo na gestão do conhecimento dentro de um contexto mais sustentável».	<i>RICI: Revista Ibero-americana de Ciência da Informação</i> , 15(2), maio-ago. 2022. ISSN 1983-5213.
«Intermediação Informacional entre as Dimensões do Desenvolvimento Sustentável» (Geraldo e Pinto, 2021).	«Considera-se importante visualizar a Sustentabilidade Informacional como uma ponte de Intermediação informacional entre as dimensões ambiental, econômica e social do Desenvolvimento Sustentável. Validando a relevância de introduzir nos estudos da Ciência da Informação, como meio de sensibilizar conscientizar e mobilizar, tanto a sociedade científica, quando a sociedade civil sobre práticas, objetivos e metas sustentáveis».	<i>XXI Encontro Nacional de Pesquisa em Ciência da Informação</i> • ENANCIB 2021 Rio de Janeiro • 25 a 29 de outubro de 2021.
«Percurso da Ciência da Informação e os Objetivos do Desenvolvimento Sustentável da Agenda 2030/ONU» (Geraldo e Pinto, 2019).	«A Ciência da Informação sendo uma área que estuda, administra e organiza meios de disseminação e uso da informação, deve estar inserida a causa da sustentabilidade, buscando suprir as necessidades organizacionais e informacionais de uma sociedade sustentável».	<i>Revista ACB: Biblioteconomia em Santa Catarina</i> , 24(2), 373-389, abr./jun., 2019.
«Informação e Desenvolvimento Sustentável: Novas Questões para o Século XXI» (Albagli, 1995).	«A viabilização de um modelo de desenvolvimento sustentável em escala global demanda, portanto, um sistema de informações qualitativa e quantitativamente distinto do atual, cuja concepção e estruturação envolve esforços dos profissionais atuantes no setor, ao mesmo tempo em que requer vontade política dos diferentes atores comprometidos com tal projeto de sociedade».	<i>Ciência da Informação</i> , 24(1), 119-125, 1995. [Artigos]

Título	Resumo/conclusão	Fonte
«Questão Ambiental e a Ciência da Informação: Estudo da Produção Científica do ENANCIB, 2015-2021» (Silva, Silva e Oliveira, 2022).	«A Ciência da informação (CI) contribui com os estudos para desenvolvimento sustentável, permitindo entender desde a produção científica, como formas de contribuir com o alcance de objetivos sustentáveis nos espaços da sociedade».	<i>Anais 29.º Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação</i> , 1(1), 1-13, 2015.
«A Sustentabilidade Informacional Pode Ser Vista Como um Novo Paradigma da Ciência da Informação?» (Geraldo, Pinto e Duarte, 2022).	«A Sustentabilidade Informacional visa utilizar o uso e o acesso à informação para alcançar o desenvolvimento sustentável, tanto para o contexto atual e para a preservação dos recursos naturais, da vida humana, da diminuição das desigualdades sociais, como buscando garantir prosperidade para as gerações futuras».	<i>Informação &amp; Informação</i> , 27(4), 229-253, out./dez. 2022.
«Análise da Produção Científica Referente à Sustentabilidade em Pesquisas de Administração» (De Luca, Cardoso, Vasconcelos e Pontes, 2014).	«Os resultados apresentados neste estudo autorizam inferir algumas tendências e perspectivas quanto às pesquisas sobre sustentabilidade, contribuindo para futuras pesquisas acadêmicas, além de servir de referência para pesquisadores que venham a se interessar pelo panorama geral da produção científica em sustentabilidade».	<i>Administração: Ensino e Pesquisa</i> , 15(3), 469-500, 2014.
«A Sustentabilidade enquanto Campo de Conhecimento: Uma Análise da Produção Científica a partir dos Clássicos Critérios de Demarcação» (Melo, 2016).	«Percebe-se que a pesquisa científica na área de Sustentabilidade não prioriza os critérios estabelecidos na Filosofia da Ciência, considerando seus autores clássicos, por praticarem uma dinâmica diferenciada».	<i>Anais do V Simpósio Internacional de Gestão de Projetos, Inovação e Sustentabilidade</i> , São Paulo, 20-22 novembro 2016.
«Consumo Sustentável: Evolução da Produção Científica e Agenda de Pesquisa» (Costa Filho, Oliveira, Lima e Silva, 2021).	«A análise da evolução histórica da pesquisa sobre consumo sustentável evidenciou que a temática manifesta um comportamento emergente, o que foi corroborado pela predominância de abordagens de pesquisa qualitativas e ensaios teóricos. Isto sugere uma oportunidade para a intensificação dos estudos no campo, inclusive aqueles de natureza quantitativa, à medida que o conhecimento sobre o fenômeno vai se consolidando no meio acadêmico».	<i>REUNIR: Revista de Administração, Ciências Contábeis e Sustentabilidade</i> , 11(1), 56-67, 2021. ISBN: 2237-3667.

Título	Resumo/conclusão	Fonte
«Análise da produção científica brasileira sobre sustentabilidade entre os anos 2008 a 2011» (Ilzuka e Peçanha, 2014).	«Analisar a produção científica em teses, dissertações e artigos publicados em periódicos de alto impacto, nacionais e internacionais, é necessário, também, para que as conclusões sejam ampliadas e fortalecidas, bem como analisar o comportamento de redes sociais com o objetivo de identificar prováveis redes de pesquisas acerca do tema».	<i>GeAS: Revista de Gestão Ambiental e Sustentabilidade</i> , 3(1), jan-abril, 1-17, 2014.
«A Produção Científica Brasileira Sobre Sustentabilidade na <i>Área Contábil</i> , no Período 2014-2019» (Silva e Melo, 2021).	«Mediante o que foi percorrido sobre a sustentabilidade e sua abordagem nas esferas científica e empresarial, conclui-se que a produção científica com temas voltados para sustentabilidade está muito abaixo diante da urgência que se vive com a crescente degradação do meio ambiente. Apesar de haver uma preocupação mundial crescente em conter os avanços da destruição pelas ações humanas».	Repositório Institucional da Universidade Federal de Alagoas.
«Sustentabilidade: A Produção Científica Brasileira Entre os Anos de 2000 e 2009» (Junqueira, Pinheiro e Maior, 2012).	«A análise desses dados revela que o comportamento da produção científica analisada confirma as premissas do presente estudo, demonstrando que a sustentabilidade ainda desperta interesse comedido dos autores nacionais».	<i>Revista Científica Hermes</i> , 6, 43-65, 2012.
«Democratização da Ciência: Uma Política Pública Necessária para o Desenvolvimento Sustentável» (Pereira e Beltrão, 2007).	«Ao longo deste trabalho buscou-se afirmar que as ações implementadas fazem parte de uma política pública de democratização do acesso ao conhecimento, de popularização da ciência, de desenvolvimento socioeconômico e cultural das populações beneficiadas. Essas ações devem ser ampliadas pelas instituições públicas de pesquisa, pois o País anseia por um processo de desenvolvimento com equidade, por políticas de inclusão social, de difusão do conhecimento e de desenvolvimento sustentável».	<i>II Encontro da Rede de Estudos Rurais</i> , 2007. Rio de Janeiro. [Tecendo o intercâmbio: o desafio do conhecimento sobre o mundo rural].
«Sustentabilidade: Reflexões Sobre Uso do Termo e Evolução de Conceitos» (Dalmago, 2021).	«Considerando que a sustentabilidade é algo em evolução, com o vetor tempo apontando para futuro, é possível especular que a simbologia numérica possa ser substituída para outra forma de avaliar a sustentabilidade».	Infoteca-e: Repositório de Informação Tecnológica da Embrapa.

Título	Resumo/conclusão	Fonte
«Produção Científica Sobre Tecnologia da Informação Verde: Estudo a Partir de Periódicos Nacionais» (Valença, Oliveira, Albuquerque Junior e Melo, 2018).	«... há pesquisas científicas a respeito do tema, porém, em quantidade extremamente inferior à necessidade atual. A busca pela sustentabilidade ambiental no uso da Tecnologia da Informação é tema recorrente na sociedade vigente, vem sendo fortemente debatido, porém, os esforços para atender esta demanda ainda são escassos. O fato impulsionador está mais ligado à redução de custos, atender questões legais e ao chamar atenção do cliente do que o cuidado com o meio ambiente em si».	<i>Interfaces Científicas: Exatas e Tecnológicas</i> , 2(3), 85-96, 13 mar. 2018.

Análise: essas produções científicas, em maioria, artigos acadêmicos, abordam sobre os desafios que enfrentam as organizações contemporâneas no gerenciamento do conhecimento produzido, seja no ambiente interno ou externo da organização. Dada essa perspectiva, as organizações buscam investir na gestão do conhecimento no contexto da sustentabilidade. Daí a importância de perceber a Sustentabilidade Informacional como uma via de intermediação informacional através do tripé das dimensões ambiental, econômica e social do Desenvolvimento Sustentável. Nesse contexto compete a CI, como intermediário para sensibilizar, conscientizar e mobilizar, seja a sociedade científica ou civil sobre práticas, objetivos e propósitos sustentáveis.

Na visão de que a CI é uma área que estuda, conduz e organiza recursos de disseminação e uso da informação, imputa-se a ela a inserção do fundamento da sustentabilidade aspirando suprir as necessidades informacionais e organizacionais de uma sociedade sustentável. Nessa visão se insere o objetivo da pesquisa realizada.

Desse modo, a CI colabora com os estudos para o desenvolvimento sustentável, assegurando o entendimento desde a produção científica como maneira de colaborar com o atingimento de objetivos sustentáveis nas esferas sociais, entende-se portanto, que a Sustentabilidade Informacional aspira utilizar o uso e o acesso à informação para alcançar o desenvolvimento sustentável, tanto no contexto atual, bem como para a preservação dos recursos naturais, da vida humana, da diminuição das desigualdades sociais, ademais, a busca de garantir prosperidade às gerações vindouras.

Foram localizadas produções científicas sobre o tema, muito embora, em quantidade extremamente inferior à necessidade emergente. Em realidade foram encontradas mais produções científicas em outras áreas, como na Administração, Contabilidade, Educação, Saúde, Arquitetura, Desenvolvimento Urbanismo, dentre outras áreas.

Conforme as produções científicas que abordam a questão da sustentabilidade, ficou claro que a busca pela sustentabilidade ambiental no uso da Tecnologia da Informação é um tema recorrente na sociedade atual, sem omitir que vem sendo fortemente debatido, porém, os esforços para atender esta demanda ainda são escassos. O fato impulsionador está mais direcionado à redução de custos, atender questões legais do que o cuidado com o próprio meio ambiente,

fator que instiga o aumento dos estudos e da elaboração de produções científicas na área da CI e a concretização no meio acadêmico.

Os resultados apresentados nessa pesquisa permitem afirmar algumas tendências e perspectivas quanto às pesquisas sobre sustentabilidade área da CI, contribuindo para futuras pesquisas acadêmicas, além de servir de referência para pesquisadores que venham a se interessar pelo panorama geral da produção científica em sustentabilidade para a informação.

A seguir foi realizada a pesquisa na Biblioteca Digital do Laboratório de Informação para Sustentabilidade (BDLIS), na qual foram visitadas 27 páginas referentes às produções científicas brasileiras, representadas nos tipos de documentos de artigo, monografia, dissertação e tese, referentes ao período 2007 a 2022, atingindo um total de 22 documentos, sendo que: 6 dissertações, 2 artigos, 8 monografias e 6 teses apresentando em seus assuntos, os seguintes termos descritores: Responsabilidade social; Responsabilidade empresarial; Sustentabilidade de social; Economia ecológica; Educação ambiental; Saúde; Desenvolvimento sustentável; Indicadores socioambientais; Gestão ambiental; Espaço urbano; Qualidade de vida; Desenvolvimento sustentável; Desenvolvimento sustentável; Meio ambiente; Recursos hídricos; Educação ambiental; Escolas sustentáveis; Conservação ambiental; Políticas públicas ambientais; Desenvolvimento sustentável; Recursos naturais; Economia ambiental; Sustentabilidade econômica; Sustentabilidade urbana; Educação ambiental; Conflito socioambiental; Insustentabilidade urbana; Administração; Recursos naturais; Economia ambiental; Preservação ambiental; Meio ambiente; Proteção ambiental, conforme mostra o Quadro 2.

Quadro 2. Levantamento de produções científicas brasileiras – BDLIS

Tipo documento	Título	Ano
Dissertação	<i>Responsabilidade Social Empresarial: Parceria Interinstitucional Trilhando Caminhos para a Sustentabilidade Social</i> (Schmidt, 2007)	2007
Artigo	«Uma Tentativa de Caracterização da Economia Ecológica» (Cavalcanti, 2014)	2004
Monografia	<i>Sustentabilidade e Conhecimento Legal na Educação Ambiental</i> (Cezar, 2014)	2014
Artigo	«Sustentabilidade da Atenção Primária à Saúde em Territórios Rurais Remotos na Amazônia Fluvial: Organização, Estratégias e Desafios» (Fausto, Giovanella, Lima, Cabral e Seidl, 2022)	2022
Tese	<i>Modelo de Avaliação da Sustentabilidade Sócio Ambiental</i> (Saldanha, 2007)	2007
Monografia	<i>Sistema de Gestão Ambiental (SGA) e a ISO 14001: Um Estudo de Caso</i> (Moreira, 2004)	2004
Tese	<i>Diálogo da Cidade de Sorocaba com a Sustentabilidade: Avaliação dos Critérios Indicativos de uma Cidade Verde e Sustentável</i> (Barroso, 2018)	2018
Dissertação	<i>Programa Nacional Escolas Sustentáveis: O Fluxo de uma Ideia no Campo das Políticas Públicas de Educação Ambiental</i> (Bianchi, 2016)	2016
Tese	<i>Aceitabilidade Social e Responsabilização pelos Impactos à Saúde: O Caso do Plano de Desenvolvimento Regional Sustentável do Xingu no Contexto da UHE Belo Monte</i> (Felipe, 2020)	2020

Tipo documento	Título	Ano
Tese	<i>Políticas Públicas de Inclusão Social e de Incentivo à Conservação do Meio Ambiente: Propostas para a (re) adaptação do Bolsa Verde como Forma de Efetivação da Sustentabilidade Socioeconômica Ambiental</i> (Assis, 2020)	2020
Tese	<i>Políticas de Áreas Protegidas como Ferramentas de Desenvolvimento: Avaliação de Impacto Fiscal dos Parques em Nível Municipal</i> (Videira, 2024)	2024
Monografia	<i>Economia do Meio Ambiente e a Valoração dos Recursos Naturais</i> (Machado, 2017)	2017
Dissertação	<i>Contribuições para uma Maior Sustentabilidade de Reassentamentos Habitacionais de Interesse Social e Ambiental: O Caso do Sítio dos Açudes, Alvorada/RS</i> (Barroso, 2004)	2004
Artigo	«Educação Ambiental e Cultura Quilombola: Uma Revisão Bibliográfica da Literatura» (Lemos, Xavier, Amorim, Lemos e Lima, 2023)	2023
Dissertação	<i>Sustentabilidade Urbana: Conceituação e Aplicabilidade</i> (Souza, 2016)	2016
TCC	<i>Sustentabilidade e Desenvolvimento Econômico: Um Estudo de Possibilidades em Direção a Uma Economia Sustentável</i> (Lino, 2011)	2011

Na sequência foram identificadas na base de dados da UFSC, 5 (cinco) dissertações e 1 (uma) tese que contextualizam a produção científica nacional em sustentabilidade como um desafio entre meio ambiente e a sociedade, e que apontam a contribuição dessas, como ainda reduzida, destacando como característica, sua distribuição geográfica mais uniforme se comparada com outras publicações nacionais.

### Quadro 3. Levantamento de produções científicas brasileiras – UFSC

Documento	Título	Ano
Dissertação	<i>Modelo Conceitual para Integração da Capacidade Absortiva e Inovação Aberta ao Contexto da Moda Sustentável</i> (Cardoso, 2024)	2024
Dissertação	<i>Arquivos Públicos Brasileiros e a Agenda 2030 da Organização das Nações Unidas: Proposta de Incorporação dos Objetivos de Desenvolvimento Sustentável</i> (Santos, 2022)	2022
Dissertação	<i>A Gestão da Sustentabilidade dos Tribunais Regionais Federais: Alinhamento com os Objetivos de Desenvolvimento Sustentável (ODS) e a Sustentabilidade Informacional</i> (Geraldo, 2021)	2021
Dissertação	<i>A Produção Científica em Sustentabilidade e a Contribuição da Ciência dos Materiais</i> (Oliveira, 2012)	2012
Tese	<i>Gestão da Sustentabilidade do Espaço Museológico Implantado no Patrimônio Edificado Mediada pela Gestão do Conhecimento: Um Estudo a partir do Museu Histórico Casa Padre Toledo em Tiradentes</i> (Baracho, 2018)	2018
Dissertação	<i>Capacidades Dinâmicas no Setor de Tecnologia da Informação: Proposição de um Modelo de Gestão em Sustentabilidade</i> (Pirett, 2019)	2019

Por último, a pesquisa foi realizada na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), biblioteca desenvolvida e coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) que integra os sistemas de informações de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil e também estimula o registro e a publicação de teses e dissertações eletrônicas. A BDTD utiliza o padrão de metadados MTD-BR, foram encontrados 740 registros de produções científicas no contexto da sustentabilidade na Ciência da Informação.

Quadro 4. Levantamento de produções científicas brasileiras – BDTD

Tipo doc.	Título	Ano
Dissertação	<i>Modelo Conceitual para Integração da Capacidade Absortiva e Inovação Aberta ao Contexto da Moda Sustentável</i> (Cardoso, 2024)	2024
Dissertação	<i>Apoio à Decisão Gerencial Suportada por Princípios de Sustentabilidade</i> (Silva, 2019)	2019
Dissertação	<i>Arquivos Públicos Brasileiros e a Agenda 2030 da Organização das Nações Unidas: Proposta de Incorporação dos Objetivos de Desenvolvimento Sustentável</i> (Santos, 2022)	2022
Dissertação	<i>A Gestão da Sustentabilidade dos Tribunais Regionais Federais: Alinhamento com os Objetivos de Desenvolvimento Sustentável (ODS) e a Sustentabilidade Informacional</i> (Geraldo, 2021)	2021
Dissertação	<i>A Produção Científica em Sustentabilidade e a Contribuição da Ciência dos Materiais</i> (Oliveira, 2012)	2012
Tese	<i>Gestão da Sustentabilidade do Espaço Museológico Implantado no Patrimônio Edificado Mediada pela Gestão do Conhecimento: Um Estudo a partir do Museu Histórico Casa Padre Toledo em Tiradentes</i> (Baracho, 2018)	2018
Dissertação	<i>Capacidades Dinâmicas no Setor de Tecnologia da Informação: Proposição de um Modelo de Gestão em Sustentabilidade</i> (Pirett, 2019)	2019

Dada a imensa quantidade de documentos visualizados na BDTD e levando-se em conta o objetivo da pesquisa, optou-se por selecionar aqueles que estivessem mais focados na área da CI. Esses documentos reportam-se à sustentabilidade nas áreas de ciência da informação, administração, gestão empresarial, moda, inovação, processo decisório, arquitetura sustentável, patrimônio histórico, bens imóveis, modelo de gestão, inovação tecnológica. Salienta-se que esses assuntos vêm ao encontro do papel do Ibict uma vez que em relação aos sistemas de informações visa gerar insumos como aparato ao desenvolvimento e ao aperfeiçoamento de pesquisas de informação tecnológica no âmbito do Projeto *Current Research Information System* (CRIS), em decorrência da estruturação da informação de pesquisa, por meio da organização da memória da informação tecnológica mediante a sistematização da atividade de pesquisa em sistema de informação nacional no âmbito de padrões internacionais.

A importância do CRIS decorre de sua capacidade de integrar para acesso conjunto, diferentes módulos de informação que permitem sistematizar a atividade de pesquisa desde a sua origem. Esse dá origem a resultados da pesquisa como (publicações, iniciativas, relatórios, produtos) entre outros.

No contexto do CRIS, para organização de sistemas nacionais, com foco em estruturação da informação sobre a pesquisa brasileira, evita duplicação e pesquisa em documentos referentes a determinada temática e implica numa gestão de resultados eficientes.

No entanto, as produções encontradas e avaliadas/analizadas, são recorrentes ao estudo conceitual, e enxerga-se a tentativa do desenvolvimento mais emergente nas diversas áreas de atuação da Tecnologia da Informação (TI) com a necessidade da sustentabilidade.

A seguir, algumas considerações fundamentadas nos resumos e conclusões de algumas dessas produções científicas selecionadas e analisadas e que merecem destaque, uma vez que fortalecem a temática em questão.

A sustentabilidade é um desafio da relação com o macro ambiente em que as organizações estão inseridas. Buscar vantagem competitiva, impõe mudanças na gestão organizacional, reforçando responsabilidades sistêmicas com o ambiente e a sociedade. Fontes geradoras de vantagem competitiva nascem de um conjunto de capacidades organizacionais desenvolvidas para que a instituição reorganize seus ativos. Essas capacidades foram denominadas «capacidades dinâmicas». O Setor de TI é dinâmico e desenvolve-se dentro do contexto empresarial, possibilitando a permanente atualização e integração dos negócios da empresa.

Com base nessas premissas, pode-se considerar que o Brasil tem contribuição relativamente pequena para as publicações sobre sustentabilidade, mas que vem crescendo a taxas muito mais elevadas que a média dos demais países. Uma característica da produção científica nacional em sustentabilidade é sua distribuição geográfica mais homogênea que a média das publicações nacionais.

As produções analisadas apontam para a importância da sustentabilidade enquanto tema de pesquisa, ainda que relativamente pequena, tornando-se assim contribuição para os estudos sobre sustentabilidade e a crescente contribuição do Brasil para os estudos no tema.

A segunda atividade referiu-se à avaliação das publicações localizadas e extração das informações que demonstrem o contexto da sustentabilidade do projeto «Estruturação da informação sobre a pesquisa brasileira».

Conforme avaliação, as produções científicas na sustentabilidade, paralelamente às das demais áreas, é considerada uma produção relativamente pequena, muito embora, pode-se perceber que a contribuição da área da CI vem contribuindo para os estudos da sustentabilidade, ainda que, a abordar a parte conceitual, com a participação inexorável da TI, ademais com a preservação que também está intrínseca no contexto do conteúdo informacional conforme a evolução tecnológica a considerar a preservação digital nos termos da fidedignidade e autenticidade da informação registrada; ademais, assegurar essas produções às pesquisas futuras das gerações vindouras, garantindo a história e o desenvolvimento da sociedade.

Quadro 5. Informações da sustentabilidade na Ciência da Informação

Áreas	Termos
Ciência da Informação	Sustentabilidade Competência em informação Gestão da informação
Ciência dos Materiais	Preservação ambiental Sustentabilidade Sustentabilidade empresarial Informação para tomada de decisão
Tecnologia da Informação	Desenvolvimento sustentável Inovação

A terceira atividade, referiu-se à análise dos formatos de representação da produção científica brasileira.

A pesquisa foi realizada tendo em vista as bases nas quais foram identificadas as produções científicas brasileiras no contexto da temática pesquisada. Identificou-se que a maioria das bases acessadas, utiliza o padrão de representação Metadados Dublin Core para representarem seus documentos digitais.

Quadro 6. Formatos de representação

Formatos	Análise
Dublin Core	Flexibilidade e interoperabilidade no acesso.
MTD2-BR	Interoperabilidade, padronização da representação, intercâmbio das produções entre as instituições brasileiras.

A quarta atividade referiu-se à indicação das produções intelectuais que podem ser inseridas no ambiente da sustentabilidade.

Essa atividade foi realizada tendo em vista as produções científicas localizadas nas bases de dados, a princípio uma atividade exaustiva, na recuperação pelas palavras - chave: Ciência da Informação – Sustentabilidade; passo que trouxe o conhecimento de um grande número de documentos que trazem em seu escopo o assunto sustentabilidade, porém, não só na área da CI, mas, em Ciência dos Materiais, Administração, Saúde, Arquitetura, Espaço urbano, Educação e moda, dentre outras áreas.

A seguir foram pinçadas as produções científicas na área da CI e voltadas especificamente para a sustentabilidade.

Em realidade, foi bem pequeno o número de produções publicadas sobre o assunto e claramente recorrentes ao despertar da consciência da importância da sustentabilidade e da necessidade emergente de maiores produções nessa questão. Vale ressaltar que em outras áreas como administração, contabilidade, educação, encontrou-se um potencial maior dessas publicações. Nesse sentido, seguem no Quadro 7 a indicação das produções intelectuais que de fato abordam o assunto de interesse dessa pesquisa.

Quadro 7. Produção intelectual

Título	Palavras-chave
«Gestão do Conhecimento e Sustentabilidade no Contexto Organizacional: Uma Revisão Sistemática na Ciência da Informação» (Fonseca, Mota e Santos Junior, 2022)	Gestão do conhecimento; Sustentabilidade; Organização; Revisão sistemática.
<i>Intermediação Informacional entre as Dimensões do Desenvolvimento Sustentável</i> (Geraldo e Pinto, 2021)	Sustentabilidade informacional; acesso à informação; Ciência da Informação; desenvolvimento sustentável
«Percurso da Ciência da Informação e os Objetivos do Desenvolvimento Sustentável da Agenda 2030/ONU» (Geraldo e Pinto, 2019)	Sustentabilidade; Desenvolvimento sustentável; Agenda 2030; Ciência da Informação.
«A Sustentabilidade Informacional Pode Ser Vista Como um Novo Paradigma da Ciência da Informação?» (Geraldo, Pinto e Duarte, 2022)	Sustentabilidade; Sustentabilidade Informacional; Ciência da Informação; Paradigma.

#### 4. CONCLUSÕES

Diante das produções científicas localizadas, lidas e analisadas, se conclui que os resultados apresentados favorecem a inferência de algumas tendências e perspectivas quanto às pesquisas sobre a informação para a sustentabilidade, no sentido de contribuir para futuras pesquisas acadêmicas, além de servir de referência para pesquisadores que venham a se interessar pelo panorama geral da produção científica da área da Ciência da Informação (CI) para a sustentabilidade.

Segundo a produção científica localizada evidenciou-se que no campo da evolução ao longo da história, especificamente no estudo sobre consumo sustentável, a questão apresenta um comportamento emergente, reforçado pelo predomínio de abordagens de pesquisa qualitativas e teorias. Perspectiva que vem abrir espaço para o aumento dos estudos na área, inclusive aqueles de natureza quantitativa, ao passo que o conhecimento sobre o fenômeno vai se consolidando no ambiente científico.

Sendo assim, entende-se que o comportamento dessas se refere às premissas da presente pesquisa, a apontar que a informação para sustentabilidade na área da CI tende a despertar interesse sistemático dos autores nacionais.

Nessa perspectiva, tenta-se consolidar que as ações implementadas em torno da sustentabilidade na CI, no contexto do CRIS apontam para elementos de uma política pública de democratização do acesso ao conhecimento, de popularização da ciência, de desenvolvimento socioeconômico e cultural das comunidades beneficiadas.

Essas ações devem ser ampliadas pelas instituições públicas, sejam de ensino ou pesquisa, visto que o Brasil aspira um processo de desenvolvimento com imparcialidade, com políticas

de inclusão social, de difusão do conhecimento e de amplo desenvolvimento sustentável essencial à evolução da área da CI na informação para a sustentabilidade.

Identificou-se que há produções científicas publicadas sobre a temática, porém, em quantidade restrita diante da necessidade emergente. A busca pela sustentabilidade ambiental no uso da Tecnologia da Informação é tema recorrente na sociedade de hoje e vem sendo fortemente debatido, porém, os esforços para atender essa demanda ainda são escassos. Atribui-se, portanto, a uma ligação maior aos altos custos a serem empregados, estar de acordo com as questões legais e ao despertar o cuidado para com o meio ambiente.

A quantidade de produções intelectuais e científicas brasileiras da CI sobre a informação para a sustentabilidade publicadas no Brasil é ainda muito restrita no contexto do reuso dos dados e informações de forma eficiente, disponibilizados em redes de pesquisas e compartilhamento na busca de identificação do potencial de inovação das produções científicas produzidas e disponibilizadas em repositórios institucionais.

Dessa maneira, identificou-se que existem muitas produções científicas replicadas nas bases, fato que vem ao encontro do objetivo do projeto em questão: enxugar, dar acesso e padronizar a considerar formato internacional.

Diante disso, ressalta-se a importância desses resultados para o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), uma vez que se trata de uma organização social e a recomendação da inserção das publicações apontadas, ainda que relativamente em pequena quantidade, que sejam indexadas, para que possam servir de aporte para instigar futuras pesquisas na temática aqui discutida.

Ressalta-se, portanto, a importância dessa pesquisa, uma vez que favorece o acesso à produção intelectual da área da CI em conexão com a sustentabilidade e; para além, suscita a produção intelectual de novos estudos que, por conseguinte, implicarão em contributo ao crescimento da produção científica na área da CI.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Albagli, S. (1995). Informação e desenvolvimento sustentável: Novas questões para o século XXI. *Ciência da Informação*, 24(1), 119-125. <https://revista.ibict.br/ciinf/article/view/617>
- Araújo, W. C. O., Inomata, D. O. e Rados, G. J. V. (2014). Desenvolvimento sustentável empresarial: O uso da gestão da informação. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 12(3), 119-135. <https://doi.org/10.20396/rdbci.v12i3.1597>
- Arruda, R. G. (2009). Unidades de informação e sustentabilidade: Requisitos para organizações do conhecimento: O caso Embrapa. *Revista Brasileira de Biblioteconomia e Documentação*, 5(1-2), 28-41. <https://rbbd.febab.org.br/rbbd/article/view/146>
- Assis, L. V. (2020). *Políticas públicas de inclusão social e de incentivo à conservação do meio ambiente: Propostas para a (re) adaptação do Bolsa Verde como forma de efetivação da sustentabilidade socioeconômica ambiental* [Tese de doutorado, Universidade Federal da Paraíba]. Repositório Institucional da UFPB. <https://repositorio.ufpb.br/jspui/handle/123456789/20879>
- Baracho, A. S. B. (2018). *Gestão da sustentabilidade do espaço museológico implantado no patrimônio edificado mediada pela gestão do conhecimento: Um estudo a partir do Museu Histórico Casa Padre Toledo em Tiradentes*

- MG [Tese de doutorado, Universidade Federal de Minas Gerais]. Repositório Institucional UFMG. <http://hdl.handle.net/1843/ECIP-B3UHSR>
- Barroso, A. M. C. (2004). *Contribuições para uma maior sustentabilidade de reassentamentos habitacionais de interesse social e ambiental: O caso do sítio dos Açudes - Alvorada/RS* [Dissertação de mestrado, Universidade Federal do Rio Grande do Sul]. Lume: Repositório Digital da UFRGS. <http://hdl.handle.net/10183/8370>
- Barroso, L. S. (2018). *Diálogo da cidade de Sorocaba com a sustentabilidade: Avaliação dos critérios indicativos de uma cidade verde e sustentável* [Tese de doutorado, Universidade Federal Fluminense].
- Cardoso, K. L. V. (2024). *Modelo conceitual para integração da capacidade absorptiva e inovação aberta no contexto da moda sustentável* [Dissertação de mestrado, Universidade Federal de Santa Catarina]. Repositório Institucional UFSC. <https://repositorio.ufsc.br/handle/123456789/254625>
- Cavalcanti, C. (2004). Uma tentativa de caracterização da economia ecológica. *Ambiente & Sociedade*, 7(1), 149-156. <https://doi.org/10.1590/S1414-753X2004000100009>
- Cezar, R. C. B. (2014). *Sustentabilidade e conhecimento legal na educação ambiental* [Trabalho de Conclusão de Curso, Universidade de Brasília]. BDM: Biblioteca Digital da Produção Intelectual Discente da Universidade de Brasília. <https://bdm.unb.br/handle/10483/9467>
- Costa Filho, F. C., Oliveira, L. V. C., Lima, D. S. V. R. e Silva, C. R. M. (2021). Consumo sustentável: Evolução da produção científica e agenda de pesquisa. *REUNIR: Revista de Administração, Ciências Contábeis e Sustentabilidade*, 11(1), 56-67. <https://reunir.revistas.ufcg.edu.br/index.php/uacc/article/view/1061>
- Dalmago, G. A. (2021). *Sustentabilidade: Reflexões sobre uso do termo e evolução de conceitos* (Documentos 193). Embrapa. Infoteca-e: Repositório de Informação Tecnológica da Embrapa. <http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1131564>
- De Luca, M. M. M., Cardoso, V. I. C., Vasconcelos, A. C. e Pontes, A. B. (2014). Análise da produção científica referente à temática de sustentabilidade em pesquisas da administração. *Administração: Ensino e Pesquisa*, 15(3), 469-500. <https://doi.org/10.13058/raep.2014.v15n3.10>
- Fausto, M. C. R., Giovanella, L., Lima, J. G., Cabral, L. M. S. e Seidl, H. (2022). Sustentabilidade da Atenção Primária à Saúde em territórios rurais remotos na Amazônia fluvial: Organização, estratégias e desafios. *Ciência & Saúde Coletiva*, 27(4), 1605-1618. <https://doi.org/10.1590/1413-81232022274.01112021>
- Felipe, M. R. (2020). *Aceitabilidade social e responsabilização pelos impactos à saúde: O caso do Plano de Desenvolvimento Regional Sustentável do Xingu no contexto da UHE Belo Monte* [Tese de doutorado, Universidade Federal de Santa Catarina]. Repositório Institucional UFSC. <https://repositorio.ufsc.br/handle/123456789/216486>
- Fonseca, D. L. S., Mota, K. C. S. L. e Santos Junior, R. L. (2022). Gestão do conhecimento e sustentabilidade no contexto organizacional: Uma revisão sistemática na Ciência da Informação. *RICI: Revista Ibero-Americana de Ciência da Informação*, 15(2), 552-570. <https://doi.org/10.26512/rici.v15.n2.2022.40038>
- Geraldo, G. (2021). *A gestão de sustentabilidade dos tribunais regionais federais: Alinhamento com os Objetivos de Desenvolvimento Sustentável (ODS) e a sustentabilidade informacional* [Dissertação de mestrado, Universidade Federal de Santa Catarina]. Repositório Institucional UFSC. <https://repositorio.ufsc.br/handle/123456789/221343>
- Geraldo, G. e Pinto, M. D. S. (2019). Percursos da Ciência da Informação e os objetivos do desenvolvimento sustentável da Agenda 2030/ONU. *Revista ACB: Biblioteconomia em Santa Catarina*, 24(2), 373-389. <https://revista.acbcs.org.br/racb/article/view/1597>

- Geraldo, G. y Pinto, M. D. S. (2021a, outubro 25-29). *Intermediação informacional entre as dimensões do desenvolvimento sustentável*. XXI Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, Rio de Janeiro. <https://enancib.ancib.org/index.php/enancib/xkienancib/paper/view/58>
- Geraldo, G. e Pinto, M. D. S. (2021b). Aspectos epistemológicos da ciência da informação e a construção conceitual da sustentabilidade informacional. Em C. R. S. Barbalho, D. O. Inomata e T. B. Fernandes (Orgs.), *Sustentabilidade informacional em ecossistemas de conhecimentos*. (pp 12-23). EDUA. <http://riu.ufam.edu.br/handle/prefix/5856>
- Geraldo, G., Pinto, M. D. S. e Duarte, E. J. (2022). A sustentabilidade informacional pode ser vista como um novo paradigma da ciência da informação? *Informação & Informação*, 27(4), 229-253. <https://doi.org/10.5433/1981-8920.2022v27n4p229>
- Ilzuka, E. S. e Peçanha, R. S. (2014). Análise da produção científica brasileira sobre sustentabilidade entre os anos de 2008 a 2011. *GeAS: Revista de Gestão Ambiental e Sustentabilidade*, 3(1), 1-17. <https://periodicos.uninove.br/geas/article/view/9836>
- Junqueira, L. P., Pinheiro, F. P. e Maior, J. C. S. (2012). Sustentabilidade: A produção científica brasileira entre os anos de 2000 e 2009. *Revista Científica Hermes*, 6, 43-65. <https://www.revistahermes.com.br/index.php/hermes1/article/view/55>
- Lemos, A. B. S., Xavier, A. R., Amorim, A. V., Lemos, P. B. S. e Lima, M. V. G. (2023). Educação ambiental e cultura quilombola: Uma revisão bibliográfica da literatura. *Revista Multidisciplinar de Educação e Meio Ambiente*, 4(3). <https://doi.org/10.51189/coneamb2023/28750>
- Lino, M. A. M. (2011). *Sustentabilidade e desenvolvimento econômico: Um estudo de possibilidades em direção a uma economia sustentável* [Monografia de graduação, Universidade Federal do Rio de Janeiro]. Pantheon: Repositório Institucional da UFRJ. <http://hdl.handle.net/11422/2158>
- Lopes, A. S. P. (2017). *Subsídios para uma política de preservação digital no contexto da Administração Pública Federal* [Dissertação de mestrado, Universidade Federal de Santa Maria]. Manancial: Repositório Digital da UFSM. <http://repositorio.ufsm.br/handle/1/13981>
- Machado, U. M. (2017). *Economia do meio ambiente e a valoração dos recursos naturais* [Trabalho de Conclusão de Curso, Universidade Federal de Santa Catarina]. Repositório Institucional UFSC. <https://repositorio.ufsc.br/xmlui/handle/123456789/174331>
- Melo, O. A. (2016, Novembro 20-22). A sustentabilidade enquanto campo de conhecimento: Uma análise da produção científica a partir dos clássicos critérios de demarcação. Em *V Simpósio Internacional de Gestão de Projetos, Inovação e Sustentabilidade, São Paulo/SP*. <https://www.singep.org.br/5singep/resultado/692.pdf>
- Moreira, A. P. (2004). *Sistema de Gestão Ambiental (SGA) e a ISO 14001: Um estudo de caso* [Dissertação de mestrado, Universidade Federal de Santa Catarina]. Repositório Institucional da UFSC. <https://repositorio.ufsc.br/xmlui/handle/123456789/87494>
- Nolin, J. (2010). Sustainable information and information science. *Information Research*, 15(2), Paper 431. <https://informationr.net/ir/15-2/paper431.html>
- Oliveira, C. M. B. (2012). *A produção científica em sustentabilidade e a contribuição da Ciência dos Materiais* [Dissertação de mestrado, Universidade Federal de São Carlos]. Repositório Institucional UFSCar. <https://repositorio.ufscar.br/handle/20.500.14289/1081>
- Pereira, F. A. e Beltrão, S. L. L. (2007). Democratização da ciência: Uma política pública necessária para o desenvolvimento sustentável. Em *II Encontro da Rede de Estudos Rurais, Rio de Janeiro*. Repositório Alice: Acesso Livre à Informação Científica da Embrapa. <http://www.alice.cnptia.embrapa.br/alice/handle/doc/894394>

- Pirett, C. N. S. (2019). *Capacidades dinâmicas no setor de tecnologia da informação: Proposição de um modelo de gestão em sustentabilidade* [Dissertação de mestrado, Universidade Federal de Uberlândia]. Repositório Institucional da Universidade Federal de Uberlândia. <http://dx.doi.org/10.14393/ufu.di.2019.991>
- Saldanha, E. E. (2007). *Modelo de avaliação da sustentabilidade sócio ambiental* [Tese de doutorado, Universidade Federal de Santa Catarina]. Repositório Institucional da UFSC. <http://repositorio.ufsc.br/xmlui/handle/123456789/90499>
- Santos, L. L. (2022). *Arquivos públicos brasileiros e a Agenda 2030 da Organização das Nações Unidas: Proposta de incorporação dos Objetivos de Desenvolvimento Sustentável* [Dissertação de mestrado, Universidade Federal de Santa Catarina]. Repositório Institucional da UFSC. <https://repositorio.ufsc.br/handle/123456789/244012>
- Santos, P. S. e Cândido, A. C. (2022). Modelo de gestão da informação para negócios sociais. *Informação & Informação*, 27(2), 423-445. <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/45174>
- Schmidt, J. A. C. (2024). *Responsabilidade social empresarial: Parceria interinstitucional trilhando caminhos para a sustentabilidade social* [Dissertação de mestrado, Universidade Estadual de Londrina]. RIUEL: Repositório Institucional UEL. <https://repositorio.uel.br/handle/123456789/10891>
- Silva, D. R. e Melo, S. G. (2021). *A produção científica brasileira sobre sustentabilidade na área contábil, no período 2014-2019* [Trabalho de Conclusão de Curso, Universidade Federal de Alagoas]. RIUFAL: Repositório Institucional da Ufal. <http://www.repositorio.ufal.br/jspui/handle/123456789/9002>
- Silva, E. C. S., Silva, G. L. e Oliveira, H. V. (2022). Questão ambiental e Ciência da Informação: estudo da produção científica do ENANCIB, 2015-2021. Em *Anais 29.º Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação*, 1(1), 1-13. <https://portal.febab.org.br/cbbd2022/article/view/2626>
- Silva, R. F. (2019). *Apoio à decisão gerencial suportada por princípios de sustentabilidade* [Dissertação de mestrado, Universidade Federal do Paraná]. Acervo Digital da UFPR. <https://hdl.handle.net/1884/66334>
- Souza, C. S. (2016). *Sustentabilidade urbana: Conceitualização e aplicabilidade* [Dissertação de mestrado, Universidade Federal de São João Del-Rei]. [https://www.ufsj.edu.br/portal2-repositorio/File/ppgtds/DISSERTACOES/Cassia\\_Souza.pdf](https://www.ufsj.edu.br/portal2-repositorio/File/ppgtds/DISSERTACOES/Cassia_Souza.pdf)
- Spink, A. (1999). Information science in sustainable development and de industrialization. *Information Research*, 1(5), Paper 65, 65-81. <https://informationr.net/ir/5-1/paper65.html>
- Valença, A. K., Oliveira, R. C. R., Albuquerque Junior, A. E. e Melo, V. F. (2018). Produção científica sobre tecnologia da informação verde: Estudo a partir de periódicos nacionais. *Interfaces Científicas: Exatas e Tecnológicas*, 2(3), 85-96. <https://doi.org/10.17564/2359-4942.2018v2n3p85-96>
- Videira, J. A. M. (2024). *Políticas de Áreas Protegidas como ferramentas de desenvolvimento: avaliação de impacto fiscal dos parques em nível municipal* [Tese de doutorado, Universidade Federal do Rio de Janeiro]. Instituto de Economia da Universidade Federal do Rio de Janeiro. <https://www.ie.ufrj.br/images/IE/PPED/Teses/2024/Tese%20Videira%20Vers%C3%A3o%20Final%20Revisada.pdf>

## EN ESTA COLECCIÓN Y EN ACCESO ABIERTO

Cordón García, J. A., Gómez Díaz, R., Alonso Arévalo, J. (2013). *Documentos electrónicos y textualidades digitales. Nuevos lectores, nuevas lecturas, nuevos géneros*.

Ediciones Universidad de Salamanca.

<https://eusal.es/eusal/catalog/book/978-84-9012-298-3>

Merlo Vega, J. (ed.) (2018). *Ecosistemas de acceso abierto*.

Ediciones Universidad de Salamanca.

<https://eusal.es/eusal/catalog/book/978-84-9012-773-5>

Cabanillas García, J., Sánchez Gómez, M., del Brío Alonso, I., Sánchez Prieto, R. (2025). *Estrategias y herramientas para la revisión sistemática de documentos científicos: un enfoque práctico con Web of Science*.

Ediciones Universidad de Salamanca.

<https://doi.org/10.14201/0MU0102>

Jorente, M., San Segundo, R., Frías Montoya, J.,

Martínez-Ávila, D., Landim, L. (coords.) (2022).

*Curación digital y género en la Ciencia de la Información: acceso y preservación*.

Ediciones Universidad de Salamanca.

<https://doi.org/10.14201/0AQ0324>

Albelda Esteban, B. (2020). *Evaluación del impacto de las bibliotecas escolares en España: Aproximación desde una doble perspectiva metodológica cuantitativa y cualitativa*.

Ediciones Universidad de Salamanca.

<https://doi.org/10.14201/0VI0444>

Crémoux, F., López Izquierdo, M., Marguet, C., Richer-Rossi, F. (eds.) (2025). *La transmisión de las noticias. Las relaciones de sucesos como espacio de mezclas, transformaciones y contaminaciones (siglos XVI - XVIII)*.

Ediciones Universidad de Salamanca.

<https://eusal.es/eusal/catalog/book/978-84-1091-058-4>

Marcello y Barriada, J., Santos Pérez, J., Sanz-Hermida, J.

(2022). *Grandezas de Brasil en la Biblioteca Histórica de la Universidad de Salamanca*.

Ediciones Universidad de Salamanca.

<https://doi.org/10.14201/0BR0002>

María Carabias Torres, A. (2012). *Salamanca y la medida del tiempo*.

Ediciones Universidad de Salamanca.

<https://eusal.es/eusal/catalog/book/978-84-9012-076-7>

López-Souto, N. , Lora Márquez, C. (eds.) (2023).

*De libros y papeles: La imprenta en la España de los siglos XVIII y XIX*.

Ediciones Universidad de Salamanca.

<https://doi.org/10.14201/0AQ0354>

González Arrieta, A. (ed.) (2019).

*Pericia documental: viejos retos, nuevas soluciones*.

Ediciones Universidad de Salamanca.

<https://eusal.es/eusal/catalog/book/978-84-9012-980-7>

Esta obra colectiva tiene como finalidad abordar desde distintas perspectivas la gestión y reutilización de los datos de investigación, elemento esencial para alcanzar los postulados sobre los que se sustenta la ciencia abierta. A través de aportaciones procedentes de distintos ámbitos, se ofrece un panorama diverso, pero con varios ejes vertebradores comunes. Entre ellos figuran, por su especial interés, la relevancia de las tareas de descripción y de preservación de los datos de investigación, el protagonismo de la biblioteca universitaria y los servicios de apoyo a la investigación en instituciones de educación superior y la evaluación de los repositorios y las bases de datos que albergan este tipo de datos. En este contexto, las ciencias de la documentación ocupan un lugar privilegiado en el estudio de los desafíos que implica la gestión de la información científica en nuestros días.

