

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 HILARIO¹ y Carlos DÍAZ-REDONDO²

¹*Universidad de Salamanca. España*
anarihi@usal.es

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

²*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

#	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.

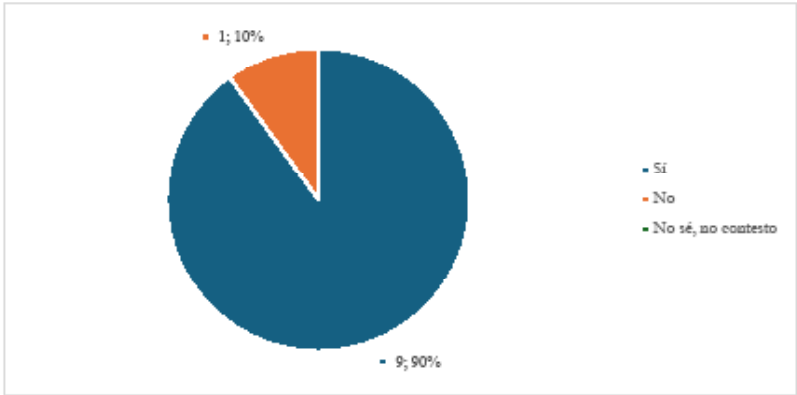
3. RESULTADOS Y DISCUSIÓN

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

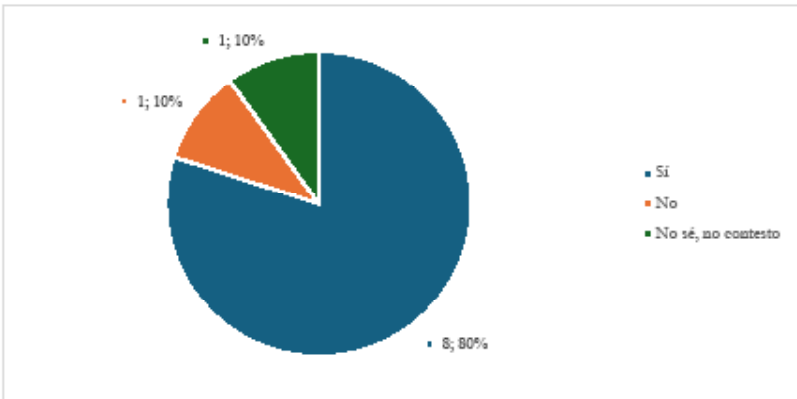


Fuente: Elaboración de los autores (2025).

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).

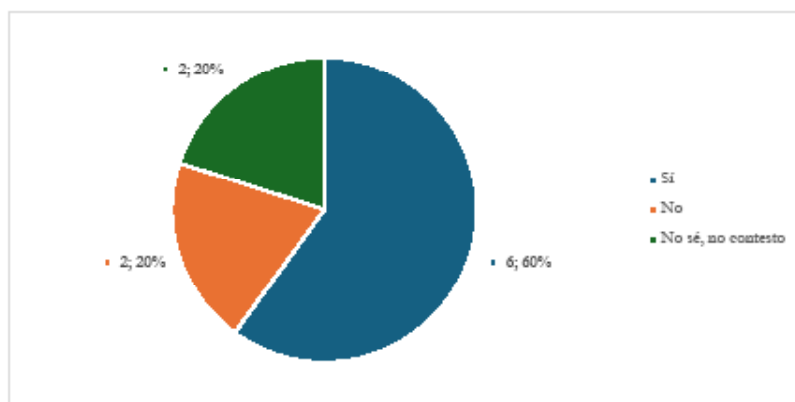
Figura 3. Indicación de los espacios de nombre en la cabecera



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



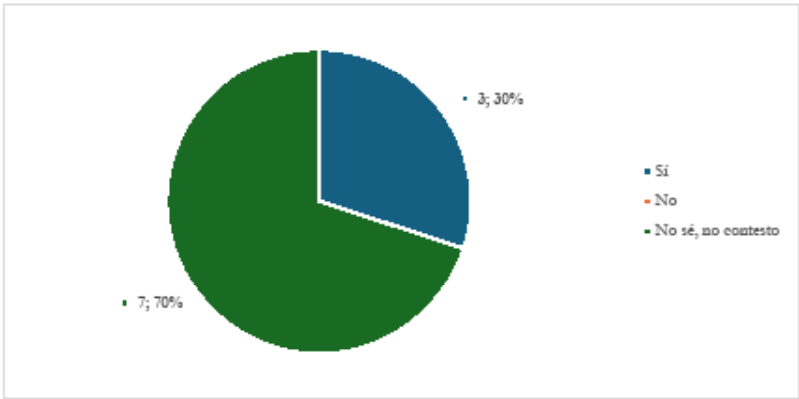
Fuente: Elaboración de los autores (2025).

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*».

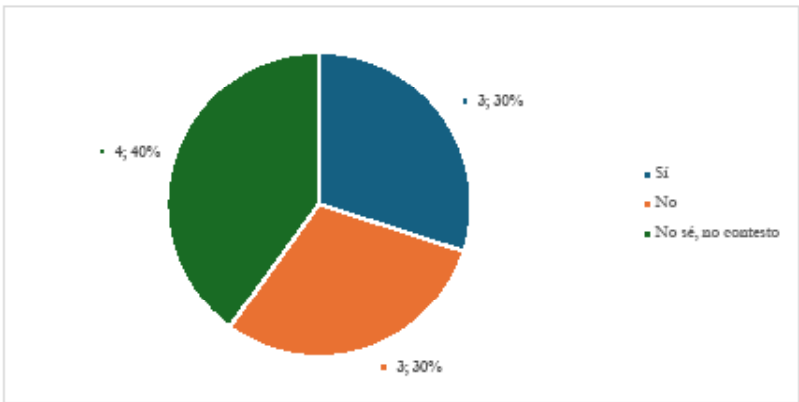
En el octavo indicador, que especifica si existe una declaración de «*Description*» en la respuesta a una orden «*Identify*», encontramos una mayor variedad de respuesta. Es así como tan solo un 30 % contesta afirmativamente. El resto del porcentaje se corresponde con el 40 % que no contesta y el 30 % 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).

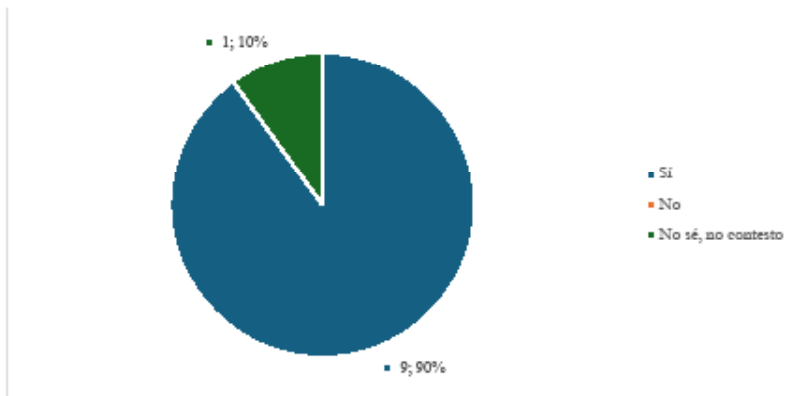
Figura 6. Declaración de «Description»



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).

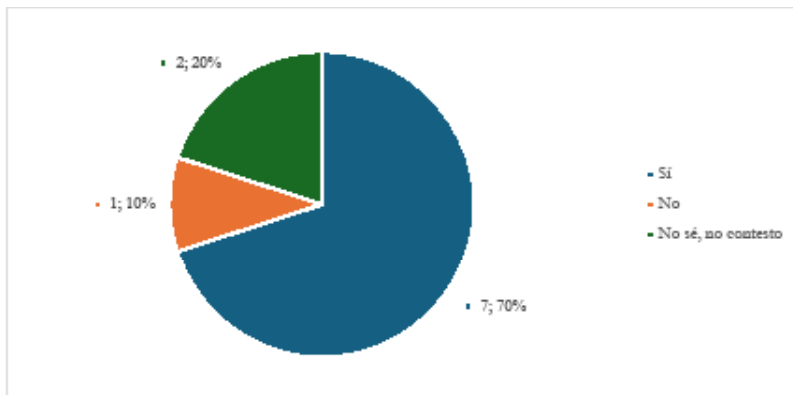
Figura 7. Entrega de registros progresiva a través de lotes



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).

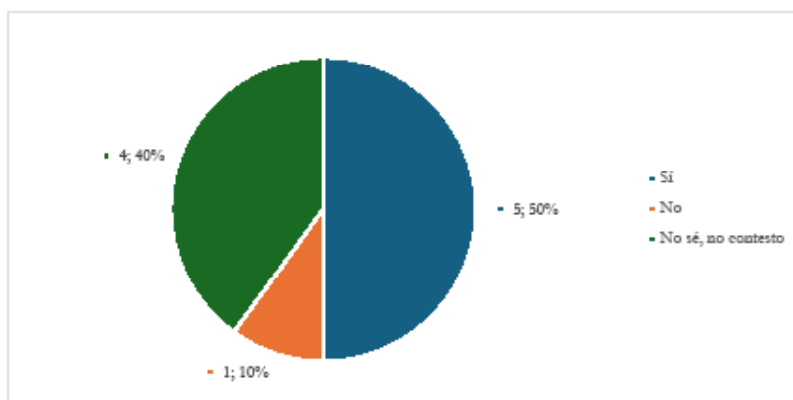
Figura 8. Tamaño de los lotes



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.

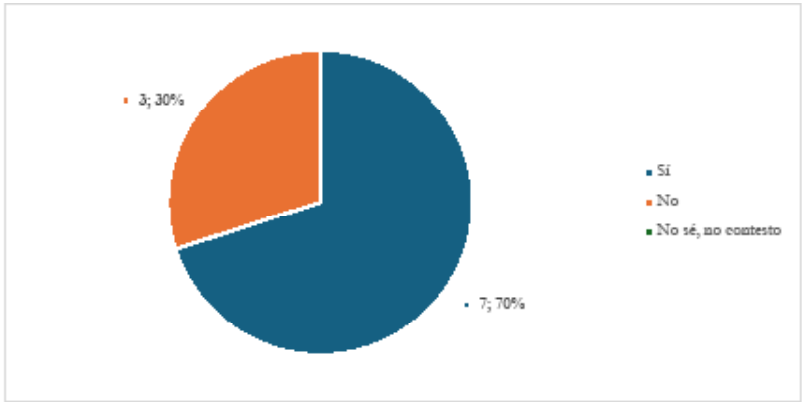
Figura 9. El formato de la fecha en la orden «*Identify*»



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).

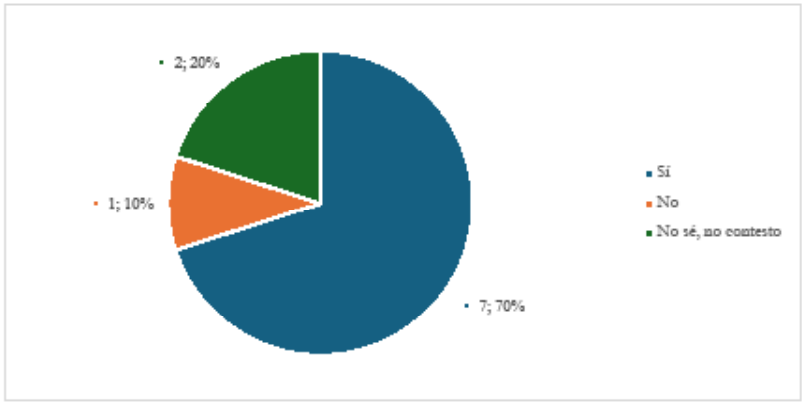
Figura 10. Integración con otros sistemas de la información



Fuente: Elaboración de los autores (2025).

El decimotercer 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).

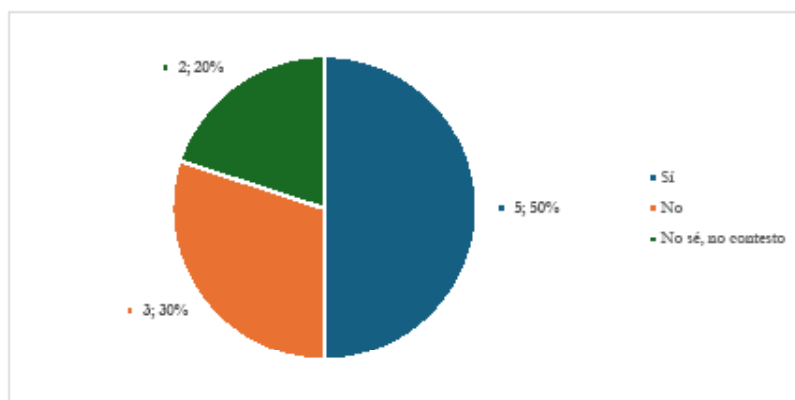
Figura 11. Inclusión de las etiquetas <meta...>



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.

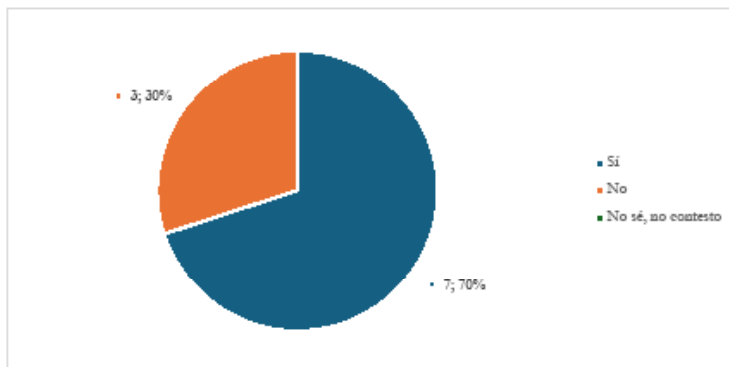
Figura 12. Implementación de Schema.org



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

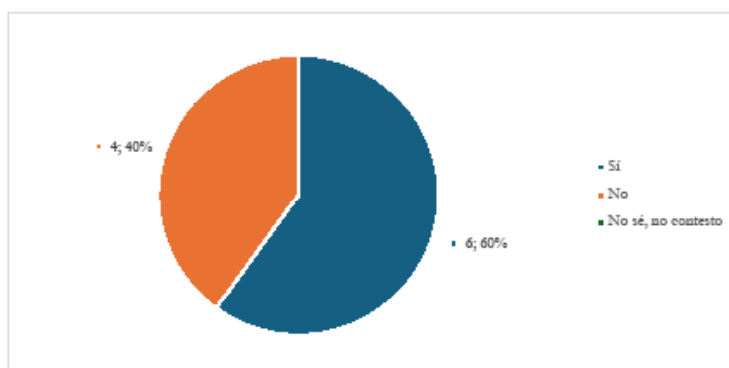


Fuente: Elaboración de los autores (2025).

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/>
- Barrueco, 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
- 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>