
ESTUDIOS / RESEARCH STUDIES

Datos abiertos de investigación en repositorios universitarios españoles

Pablo Monteagudo-Haro

Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid
Correo-e: pablo_monteagudo3@hotmail.com | ORCID iD: <https://orcid.org/0009-0003-2874-348X>

Juan Jose Prieto-Gutierrez

Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid
Correo-e: jjpg@ucm.es | ORCID iD: <https://orcid.org/0000-0002-1730-8621>

Recibido: 11-11-2023; 2ª versión: 20-12-2023; 3ª versión: 5-01-2024; Aceptado: 22-01-2024;
Fecha de publicación: 16-10-2024

Cómo citar este artículo/Citation: Monteagudo-Haro, P., Prieto-Gutierrez, 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>

Resumen: Se analiza la situación actual de los datos de investigación abiertos en los repositorios de las universidades españolas a través de doce indicadores que permiten compararlos entre sí. Los doce indicadores de elaboración propia tratan sobre los conjuntos de datos de investigación y las políticas institucionales vinculadas al acceso abierto, además de algunas de las características claves de los repositorios. La metodología que se lleva a cabo consiste en la comparación entre repositorios de las diferentes universidades vinculadas a REBIUN. El resultado ha sido que los conjuntos de datos en repositorios institucionales son escasos, y la situación es heterogénea en todo el territorio. Esto lleva a plantear cuestiones sobre cuáles van a ser las futuras políticas de acceso abierto sobre los datos de investigación en las principales instituciones científicas del país.

Palabras clave: Datos de investigación, repositorios institucionales, ciencia abierta, universidades españolas.

Open Research Data in Spanish University Repositories

Abstract: The current situation of open research data in Spanish university repositories is analyzed by means of twelve indicators that allow us to compare them with each other. The twelve self-developed indicators deal with research datasets and institutional policies linked to open access, as well as some of the key characteristics of the repositories. The methodology used consists of comparing the repositories of the different universities linked to REBIUN. The result has been that datasets in institutional repositories are scarce, and the situation is heterogeneous across the territory. This raises questions about future open access policies for research data in the country's main scientific institutions.

Keywords: Research data, academic repositories, open science, Spanish universities.

1. INTRODUCCIÓN

En 2002 se presentaron las bases del acceso abierto a través de la *Budapest Open Access Initiative* (Open Society Institute, 2002), que planteaba la retirada de las barreras a la literatura científica para acelerar la investigación. Proponía como estrategias para ello el autoarchivo institucional y las publicaciones de acceso abierto. Posteriormente, *Bethesda Statement on Open Access Publishing* (Howard Hughes Medical Institute, 2003) y *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (Max Planck Gesellschaft, 2003), ambas de 2003, buscarían profundizar el acceso abierto implicando a actores concretos más allá de los investigadores, como son las bibliotecas, los editores, las instituciones y las agencias de financiación.

La consideración de acceso abierto implica la posibilidad de leer, descargar, copiar, distribuir, imprimir, buscar o usar con cualquier propósito legal, sin ninguna barrera financiera, legal o técnica, fuera de las que son inseparables para el mismo acceso a Internet (Open Society Institute, 2002).

La Directiva (UE) 2019/1024 habla específicamente del acceso abierto en el ámbito de las universidades y de las instituciones de investigación. El *Open Access* (acceso abierto), el *Open Science* (ciencia abierta) y el *Open Data* (datos abiertos) se interrelacionan en el ámbito universitario y científico europeo.

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 reproducibilidad 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 reproducibilidad y la evaluación del proceso, pero también, permiten la reutilización científica.

Es necesario plantear una distinción entre datos abiertos y datos de investigación abiertos. La consideración de los datos abiertos está en relación con los datos que produce una institución en el desarrollo de su actividad, que en el caso de las administraciones públicas se vincula con el *Open Government* (gobierno abierto). Las iniciativas de gobierno abierto buscan un control sobre las acciones del gobierno y de las diferentes instituciones por parte de cualquier ciudadano o de cualquier

grupo social interesado. De este modo se puede considerar el gobierno abierto como el modo de plantear transparencia con la información gubernamental, apertura en la participación y colaboración ciudadana en los mecanismos de decisión y control; se trata de un modelo que se presenta dentro de los estándares de la democracia participativa (Clabo y Ramos-Vielba, 2015). En este contexto, se desarrolla la posibilidad de la reutilización de los datos abiertos de las administraciones públicas.

Un ejemplo de ello se llevó a cabo en el estudio de 2018 sobre los datos que la universidad como institución generaba en torno a los datos de gestión y a los datos académicos, siguiendo los estándares de las políticas para la reutilización de información del sector público (Martín González y Ríos Hilario, 2018). En este estudio se explicitaba que la universidad producía tres tipos de datos: los de gestión, los académicos y los datos de investigación (Martín González y Ríos Hilario, 2018: 117). Este punto de distinción es clave para diferenciar los datos abiertos, que suelen vincularse al gobierno abierto, y los datos de investigación abiertos, que se vinculan al desarrollo científico.

Los datos de investigación, sean abiertos o no, son la prueba objetiva a partir de la que se valida unas tesis o postulados de investigación. Los datos de investigación abiertos son parte de la ciencia abierta en tanto y en cuanto se plantee una política de difusión abierta. En este contexto científico y de investigación, el *Data Sharing* (datos compartidos) se plantea como "la acción de compartir con otros colegas los ficheros de datos (o raw data) generados durante el curso de una investigación" (Torres Salinas y otros, 2010: 258), es decir, compartir el material sin procesar que se ha generado en el curso de la investigación. El *Data Sharing* se vincula a la filosofía de acceso abierto porque favorece la apertura y la accesibilidad, además, promueve una "segunda vida" para esos datos sin procesar, pudiendo reutilizarlos para otro fin (Sixto-Costoya y otros, 2019).

Compartir los datos de investigación abiertos permite incrementar el impacto y la visibilidad de la investigación, mejorar la reproducibilidad de la ciencia, la posibilidad de reutilización de los mismos datos que el estudio original en otros estudios posteriores, abrir la posibilidad a un ahorro de costes, fomentar la colaboración y aumentar la credibilidad en el sistema científico (Lyon, 2016).

En cuanto a la gestión de la información científica, la concepción de la curación de datos (*Data Curation*, en inglés) ha generado una nueva área de responsabilidad para investigadores, bibliotecarios y profesionales de la información en el entorno

de las bibliotecas digitales (Heidorn, 2011), los cuales se dedican a la búsqueda, selección, caracterización y difusión continua del contenido más relevante de diversas fuentes de información (Gualar y Leiva-Aguilera, 2013: 27), y, en definitiva, de datos (Tammaro, y otros, 2019).

En el caso de la investigación en el contexto europeo y español se menciona la cuestión en torno a la transferencia de los resultados financiados mayoritariamente con fondos públicos (Ley 37/2007; Ley 14/2011; Ley 18/2015 y Directiva (UE) 2019/1024), refiriéndose tanto a las publicaciones como a los datos de investigación. El sentido de este interés por compartir los resultados de las investigaciones de modo abierto tiene como fin mejorar el retorno de la inversión que hacen las instituciones públicas cuando financian la investigación (Hernández-Pérez, 2016: 520) y garantizar el acceso abierto a los datos para el público, los gobiernos y las agencias de financiación (Stieglitz y otros, 2020). En los últimos años se ha facilitado el depósito de los datos a través de varias iniciativas, como por ejemplo los repositorios Zenodo (creado por OpenAIRE y el CERN), Figshare, Data-Cite, etc.

En el contexto de las políticas de acceso abierto de las universidades españolas, hace diez años tres cuartas partes habían iniciado alguna acción en el marco de la ciencia abierta dirigida hacia la creación de repositorios o cursos OpenCourseWare (OCW) (Abadal y otros, 2013). Los cursos OCW buscaban que los materiales educativos estuvieran en abierto y de forma gratuita para poder ser consultados. Hoy en día, el conocimiento de las políticas ha aumentado de forma generalizada hacia la forma de compartir datos (González-Teruel y otros, 2022). Además, en la actualidad, el marco regulatorio promueve el desarrollo de repositorios institucionales o temáticos de acceso abierto, propios o compartidos vinculados a las universidades y organismos de investigación españolas para transferir los resultados de las investigaciones a la sociedad (Ley 17/2022 que modifica la Ley 14/2011), que afecta principalmente a las publicaciones. Asimismo, el artículo 12 de la Ley Orgánica del Sistema universitario (2023) obliga al personal docente e investigador a hacer pública la versión final de sus publicaciones en revistas científicas mediante el depósito en un repositorio. Mientras que el mandato legal español plantea para los datos de investigación el alojamiento en un repositorio (Ley 37/2007), sin tener que ser, necesariamente, en un repositorio institucional.

En este momento, los repositorios institucionales de las universidades españolas albergan publicaciones pre-prints y post-prints, entre otros

documentos, y siguiendo este mismo sentido, se podría plantear que este mismo espacio pudiera servir para albergar los datos de investigación. Datos que, de algún modo, están vinculados a esas publicaciones que albergan los repositorios institucionales, y que se desarrollan en las instituciones académicas que los mantienen. Estos datos deberían cumplir los principios internacionales FAIR (por sus siglas en inglés, Findable, Accessible, Interoperable y Reusable), para tratar de que sean sencillos de encontrar, accesibles, interoperables y reutilizables.

La vinculación entre datos de investigación y los repositorios institucionales se plantea bajo el interés que pudiera haber entre las bibliotecas académicas y estos datos, puesto que: "Los repositorios de datos de investigación sirven, entre otros fines, para validar resultados de investigación y, por tanto, deben estar vinculados de alguna manera a las publicaciones científicas en donde se muestra para qué fueron utilizados esos datos" (Hernández-Pérez y García-Moreno, 2013: 261).

No hay un mandato legal expreso del empleo, en exclusiva, de los repositorios institucionales para albergar datos de investigación. Las universidades españolas emplean esta plataforma en su inmensa mayoría para almacenar su producción científica.

Para concretar esta finalidad se desarrollan los límites de la investigación a partir del siguiente epígrafe sobre objetivos del estudio.

2. OBJETIVOS

El objeto de estudio son los datos de investigación en abierto en las universidades españolas, específicamente su situación en los repositorios universitarios. El objetivo del estudio es analizar a través de una comparación la situación de los repositorios institucionales de las universidades de España. Para ello, se van a buscar los aspectos relevantes que permitan dibujar la situación actual, y así, poder señalar algunos pasos que podrían ser necesarios para alcanzar los objetivos que el marco legal pretende en relación con los datos abiertos.

Se precisan los siguientes objetivos específicos:

1. Analizar las características descriptivas de los repositorios institucionales de las universidades españolas que forman parte de la Red de Bibliotecas Universitarias REBIUN (Red Española de Bibliotecas Universitarias).

2. Analizar indicadores previamente establecidos para mostrar el acceso a los datos de investigación de las universidades españolas en sus repositorios, y el contexto digital en que se encuentran.

3. MÉTODO

Con el fin de dar respuesta a los objetivos planteados, se usa un método comparativo entre los datos extraídos en la observación web de los repositorios institucionales de las universidades españolas. El método comparativo consiste en analizar los repositorios de las universidades españolas estableciendo semejanzas y diferencias entre los resultados que se obtienen universidad por universidad, permitiendo comparar los resultados pudiendo tener en cuenta la titularidad (pública o privada), las comunidades autónomas, o incluso, el tipo de software del repositorio, entre otros muchos aspectos.

Para obtener los resultados se observa en la web de cada una de las instituciones una serie de indicadores de elaboración propia (Tabla I), la observación fue realizada durante el mes de julio de 2022. Estos indicadores son los siguientes:

El objetivo principal, situación actual de los datos en los repositorios, se plantea tras desarrollar los dos objetivos específicos, análisis de las características descriptivas y de los indicadores, puesto que es necesaria la recogida previa de estos resultados para poder elaborar una descripción, primero cuantitativa, de la situación de los datos en 2022.

El primer grupo de indicadores, "Datos de investigación en repositorios institucionales", que comprenden los indicadores i.01, i.02 y i.03, y el segundo grupo de indicadores, "Políticas de acceso abierto", indicadores i.04 y i.05, se aplican a toda la muestra de repositorios de las universidades vinculadas a REBIUN, mientras que el tercer grupo, "Características de los repositorios", indicadores del i.06 al i.12, se aplican a las instituciones que albergan datos en sus repositorios institucionales, o tienen una sección para albergar datos, puesto que el análisis de los repositorios es de interés en relación a los datos de investigación, y no a la evaluación en sí de los repositorios institucionales.

Tabla I. Indicadores para el estudio de datos de investigación en repositorios institucionales.

DATOS DE INVESTIGACIÓN EN REPOSITARIOS INSTITUCIONALES			
	INDICADORES	RESULTADOS	DESCRIPCIÓN
i.01	Alojamiento datos de investigación	Sí/No	Si el repositorio institucional aloja o no datos de investigación.
i.02	Ubicación de los datos	Sí/No	Si dentro del propio repositorio existe un espacio propio para alojar los datos de investigación.
i.03	Nº conjunto de datos	Nº entero	Indica la cantidad de conjuntos de datos que se encuentran alojados en el repositorio.
POLÍTICAS DE ACCESO ABIERTO			
	INDICADORES	RESULTADOS	DESCRIPCIÓN
i.04	Política Acceso Abierto General	Sí/No	Si la institución cuenta con un documento publicado de política de acceso abierto.
i.05	Gestión de datos de investigación	Sí/No	Si la institución cuenta con un documento publicado que refleje la política de gestión de datos de investigación.
CARACTERÍSTICAS DE LOS REPOSITARIOS			
	INDICADORES	RESULTADOS	DESCRIPCIÓN
i.06	Software	Nombre del programa	Se indica el tipo de software que soporta el repositorio institucional.
i.07	Tipo de metadatos	Protocolos de metadatos	Se recogen los diferentes protocolos de metadatos por institución.
i.08	Organizar por metadatos	Sí/No	Se indica si se permite organizar los conjuntos de datos en función de las distintas variables de los metadatos.
i.09	Accesibilidad	Sí/No	Se indica si se puede acceder o no a los documentos de los conjuntos de datos.
i.10	Licencias y políticas de derechos de autor	Tipos de licencias	Se recoge el tipo o los tipos de licencia de propiedad intelectual que el repositorio utiliza en los conjuntos de datos.
i.11	Recolectores	Nombres de los recolectores	Se indica a qué recolectores están vinculados los repositorios.
i.12	Guía de publicación	Sí/No	Se indica si el repositorio o la biblioteca ofrecen guías de publicación de datos de investigación.

Los indicadores de análisis se diferencian de otras propuestas, como el trabajo de REBIUN sobre repositorios, Guía para la evaluación de Repositorios institucionales de Investigación (Barrueco y otros, 2021). No se replican sus ocho criterios, puesto que cada uno de ellos se extiende en un análisis muy pormenorizado, y superaría los objetivos de esta investigación. Se atiende a este documento de REBIUN porque es el marco de referencia para las instituciones académicas españolas, pero los indicadores también se han planteado a partir del texto FAIREST: A Framework for Assessing Research Repositories (d'Aquin y otros, 2023), puesto que tiene en cuenta específicamente los datos de investigación en los repositorios. Da un análisis más próximo al objetivo de la investigación, pero, al ser desde una perspectiva internacional en algunos puntos se ha tenido que adaptar a la realidad específica española.

La muestra de análisis se centra en las 76 universidades españolas miembros de la CRUE y REBIUN, 50 de ámbito público y 26 de ámbito privado. De la muestra hay que señalar que la Universidad de Vigo y su repositorio Investigo no se han analizado ya que no se pudo acceder por problemas técnicos de la universidad durante el periodo de búsqueda de información, haciendo que la muestra final sea de 75 instituciones.

4. RESULTADOS

El primer aspecto que se va a abordar es el de los datos de investigación en los repositorios institucionales.

4.1 Datos de investigación en repositorios institucionales

El indicador "alojamiento de datos de investigación" (i.01) se toma en primer lugar porque va a dar una visión global sobre los datos en las 76 universidades de REBIUN. Se indica el almacenamiento de datos en todas las universidades, y también, se desagregan los resultados de los datos en relación con si la institución es pública o privada (Figura 1).

Las instituciones públicas lideran el almacenamiento de datos de investigación en repositorios. Si sólo atendemos al grupo de instituciones públicas el resultado que alberga datos es del 73%, que son 36 universidades, mientras que, si atendemos al total de las instituciones españolas el porcentaje baja a 52% (39 instituciones). Los repositorios de universidades públicas con datos de investigación son seis veces superior al de las universidades privadas que es del 12% (3 instituciones), siendo estas la Universidad Camilo José Cela, Mondragon Unibertsitatea y Universitat Oberta de Catalunya.

Por comunidades autónomas destaca Andalucía y Cataluña, con ocho instituciones con datos de investigación cada una, y Madrid, con siete (Figura 2).

Figura 1. Porcentaje de universidades REBIUN que guardan datos de investigación: porcentaje total, porcentaje de instituciones privadas y porcentaje de instituciones públicas.

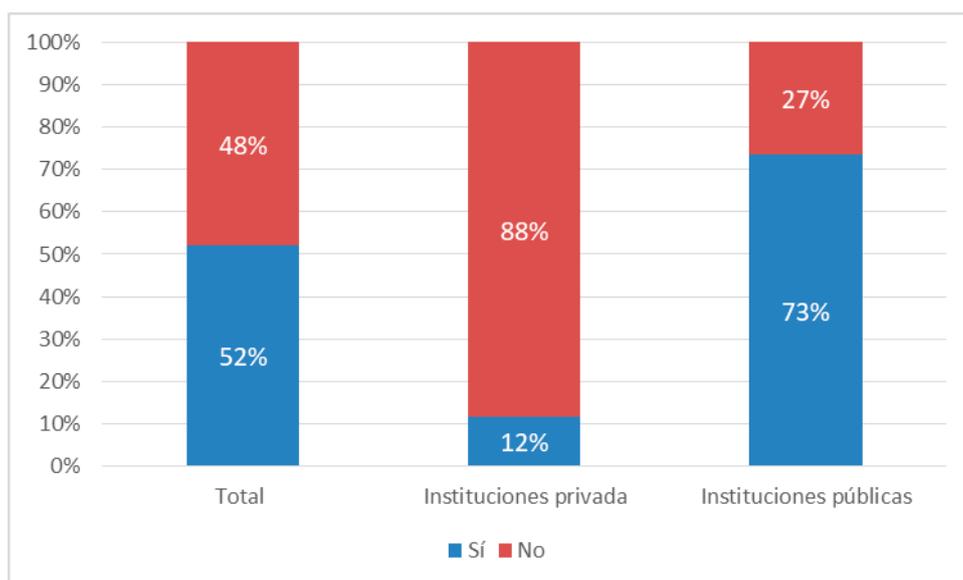


Figura 2. Número de universidades REBIUN con datos de investigación en sus repositorios por comunidades autónomas en España.

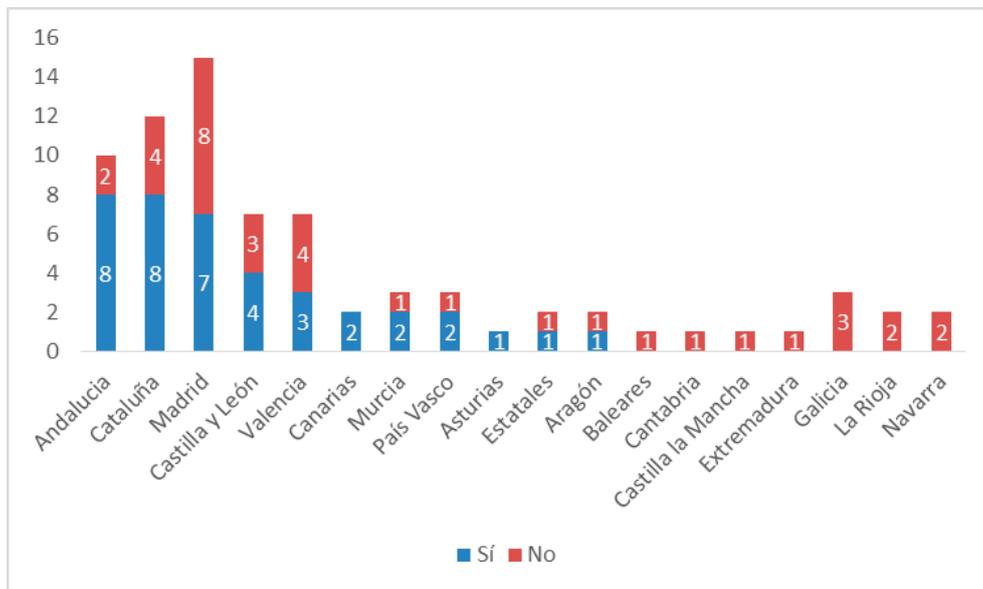
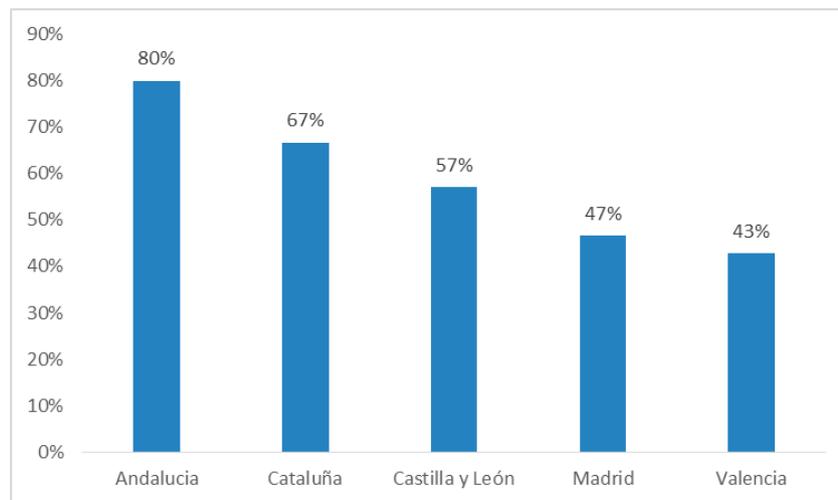


Figura 3. Porcentaje de universidades de la comunidad que tienen datos en sus repositorios.



En la Figura 3 se muestran las cinco comunidades autónomas con mayor número de universidades que tienen en sus repositorios datos de investigación, que son: Andalucía, Cataluña, Castilla y León, Madrid y Valencia. Los ocho repositorios de Andalucía representan el 80% de las instituciones de la región, mientras que en Cataluña sólo son el 67%; en Castilla y León representa el 57%, en Madrid, el 47%, y en Valencia, el 43%.

Si comparamos la Figura 2 y la Figura 3 se observa que: (a) Andalucía y Cataluña tienen el

mismo número de universidades comprometidas con la publicación de datos en sus repositorios (ocho cada una), sin embargo, analizando el porcentaje parece que la comunidad andaluza está más comprometida, dando un resultado del 80% frente al 67%; (b) mientras que Madrid pasa a la cuarta posición si atendemos al porcentaje, pero a la tercera posición si atendemos al total de universidades con datos en sus repositorios.

El segundo indicador, "ubicación de los datos" (i.02), que define si el repositorio clasifica los con-

juntos de datos de investigación en un apartado específico, permite identificar el nivel de desarrollo de la accesibilidad a los datos en los repositorios.

En la Figura 4 se puede ver que el porcentaje de instituciones que tienen un apartado de datos de investigación en sus repositorios es del 56%, un porcentaje superior al del 52% de instituciones que contienen datos de investigación (Figura 1). Esta aparente diferencia está producida por tres universidades, la Universidad Internacional de La Rioja, la Universidad Francisco de Vitoria y la Universidad Miguel Hernández de Elche, que han creado secciones específicas para datos de investigación en sus repositorios, pero no contienen ningún dato en julio de 2022.

El indicador "número de conjunto de datos" (i.03) cuantifica el número de conjuntos de datos que alberga cada repositorio.

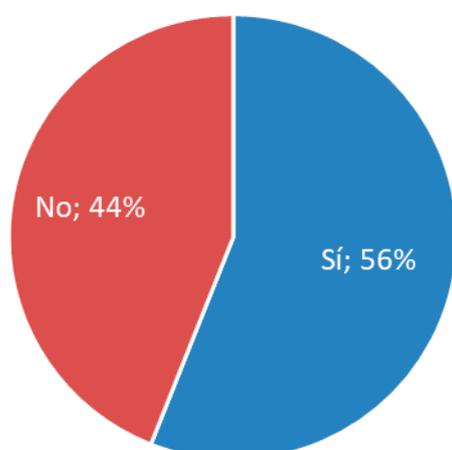
Hay que señalar algunos patrones que han surgido al revisar los registros:

(1) los datos de investigación son publicados por autores recurrentes, el caso más representativo es el de la Universidad de Zaragoza, donde todos los conjuntos de datos pertenecen a una misma persona;

(2) en las secciones creadas para almacenar datos de investigación en algunos casos se encuentran otro tipo de registros, como documentos de apoyo a la investigación que no son propiamente datos de investigación.

En los datos que presenta la Figura 5 se puede ver que solo hay cuatro universidades que tienen almacenados más de 200 datos de investigación: la

Figura 4. Porcentaje de universidades con una sección específica de datos de investigación.



Universidad de Alcalá (375), la Universitat Pompeu Fabra (268), la Universidad Carlos III de Madrid (243), y la Universidad de Zaragoza (222). Destaca que entre estas cuatro universidades hay dos de la comunidad de Madrid.

La Figura 6 nos presenta los conjuntos de datos cruzados por comunidades autónomas, donde podemos ver que Madrid, Cataluña y Andalucía se encuentran en los primeros puestos.

En la Figura 2 podíamos ver como Andalucía, Cataluña y Madrid contaban con un número muy parecido de repositorios que contenían datos de investigación: ocho en Andalucía, ocho en Cataluña, y siete en Madrid. Y, por otro lado, en la Figura 3 se representaba el porcentaje de Universidades REBIUN con datos de investigación en sus repositorios por comunidades autónomas: Andalucía lideraba la Figura con un 80%, Cataluña, con un 67% era la segunda, y Madrid, con un 47% pasaba a la cuarta posición. Si contraponemos estos datos al número de conjunto de datos por comunidad autónoma se puede ver que Madrid lidera el número de datos en repositorios, seguido de Cataluña y de Andalucía, casi triplicando el número de datos de Madrid a los datos alojados en los repositorios en Andalucía. Aquí se muestra que una comunidad autónoma para alojar en sus repositorios datos de investigación no necesita tener un mayor número de repositorios en sí, pero sí que se ve una correlación entre las comunidades autónomas que mayor número de repositorios tienen y la cantidad de datos almacenados.

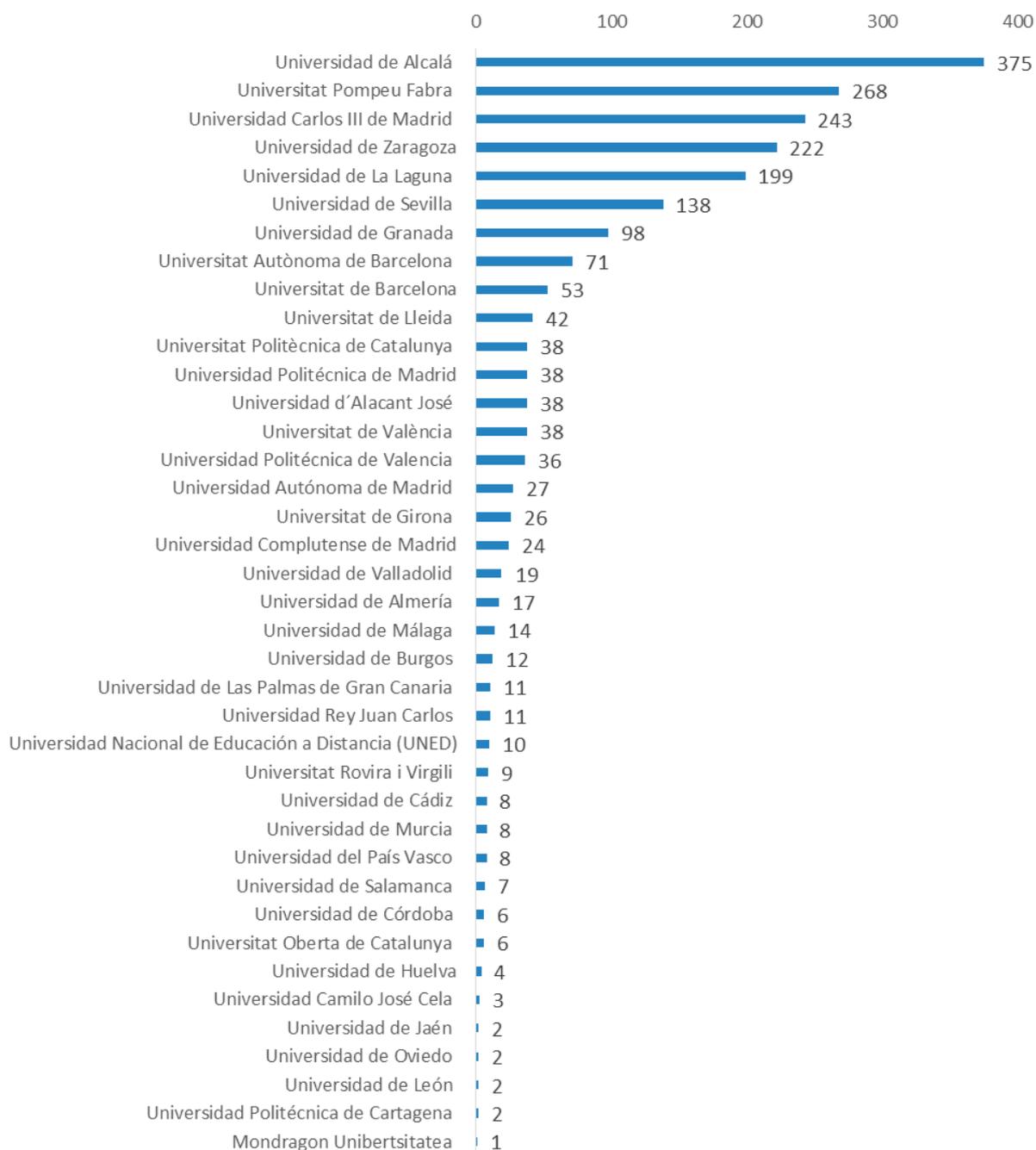
En Madrid el 36% de los datos totales están producidos dentro del Consorcio Madroño, conformado por la Universidad Rey Juan Carlos, la Universidad Autónoma de Madrid, la Universidad Carlos III de Madrid, la Universidad Politécnica de Madrid, la Universidad de Alcalá y la UNED. Estas instituciones representan un 9% del global de instituciones de REBIUN.

4.2 Políticas de acceso abierto

En este subapartado se plantea la vinculación del repositorio con el acceso abierto, para ello se observan dos indicadores: si la institución tiene una política de acceso abierto en general declarada (i.04), y si tiene una política específica de la gestión de los datos de investigación (i.05).

De las 75 instituciones analizadas el 65% (49 instituciones) tienen accesible una política de acceso abierto (Figura 7). Se coteja si alojan un documento público que indica la posición de la institución sobre su accesibilidad y su autopublicación, es decir, se ha hecho una comprobación tanto en el repositorio como en la página web general

Figura 5. Nº conjuntos de datos de investigación en el repositorio instiucional por universidad.



de la universidad sobre si tiene este documento de accesibilidad, no solo en el repositorio.

El indicador "gestión de datos de investigación" (i.05) analiza los documentos de las políticas de gestión de datos de la institución para datos de investigación, permite ver el grado de desarrollo del protocolo institucional para el tratamiento más

adecuado de los datos de investigación para cada institución.

En la Figura 8 se puede ver que solo el 9% (siete instituciones) de las universidades analizadas tienen una política específica para los datos de investigación.

Figura 6. Nº de datos de investigación en repositorios institucionales por comunidad autónoma.

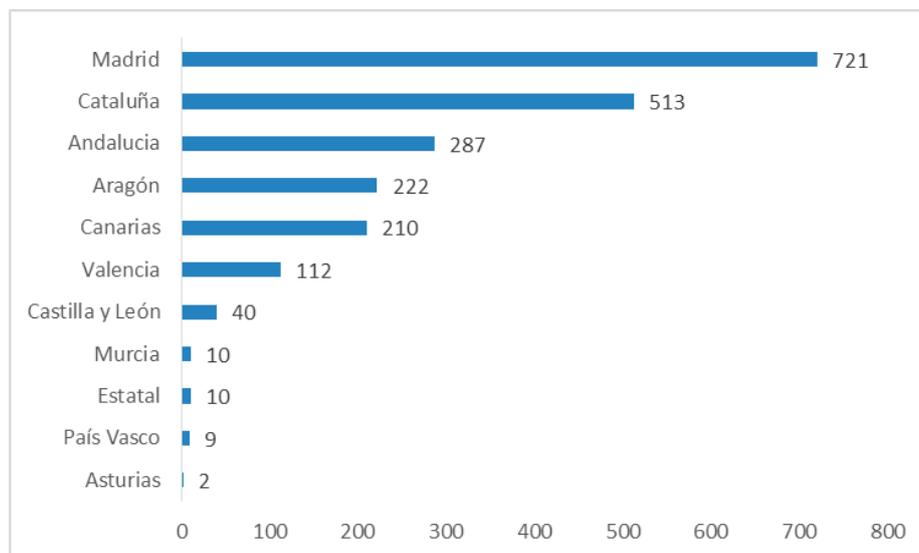


Figura 7. Instituciones REBIUN que tienen una Política de Acceso Abierto.

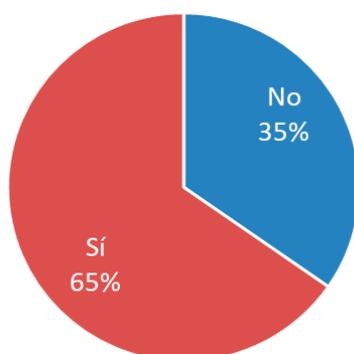
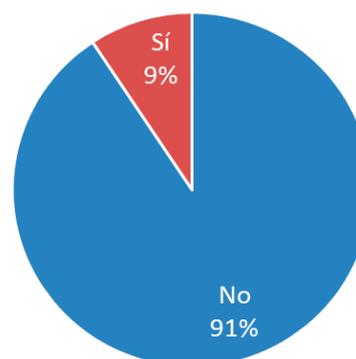


Figura 8. Instituciones REBIUN con Política de Gestión de Datos de Investigación.



En la Figura 8 también se aprecia que el 91% (sesenta y ocho instituciones) no tienen una política de gestión de datos ni una sección específica sobre esta gestión dentro de sus políticas de acceso abierto. Dentro de este grupo, no obstante, hay una división interna entre las organizaciones que no citan en sus políticas de acceso abierto los datos de investigación, cincuenta y ocho de sesenta y ocho universidades, y las que hablan de los datos de investigación como una parte de la producción científica junto a los artículos y otros trabajos científicos, diez de las sesenta y ocho instituciones.

Seis de las siete instituciones que tienen accesible este documento de gestión de datos de investigación forman parte del Consorcio Madroño; la otra

universidad es la Universidad de Málaga. La situación de colaboración del Consorcio Madroño no se repite en ninguna otra comunidad autónoma. Podría ser considerada una tercera forma de compartir los datos de investigación que se suma a las dos opciones mencionadas, que son: o bien, alojar los datos en su repositorio, o bien, en un repositorio externo.

4.3 Características de los repositorios

En este apartado se desarrollan los indicadores que permiten analizar la situación actual de los repositorios institucionales, dando pie a analizar el contexto de los repositorios universitarios donde se observa el alojamiento de los datos de investigación.

De los setenta y cinco repositorios activos en julio de 2022, cuarenta y dos tienen un apartado para datos de investigación, y treinta y nueve contienen algún tipo de dato. El grupo de repositorios que se analizarán en este apartado son los cuarenta y dos repositorios.

El indicador sobre software (i.06) identifica el programa que soporta el repositorio, siendo DSpace el software que predomina con un 76,19% (32 repositorios), seguido de Dataverse, 14,29% (6 repositorios), CDS Invenio es usado por dos repositorios, y finalmente, Fedora y E-Prints es usado en un único caso cada uno.

Sobre metadatos (i.07), se observa que todas las instituciones con datos de investigación usan Dublin Core (treinta y nueve instituciones), mientras que seis utilizan METS, cinco instituciones usan MARC, y el mismo número de instituciones usan PREMIS, mientras solamente cuatro usan MODS, cuatro EDM, y dos MARCXML (Figura 9).

El 76% de repositorios institucionales con apartado para datos de investigación solo usan un único estándar de metadatos, que es Dublin Core.

En cuanto a la organización por metadatos (i.08) de las cuarenta y dos instituciones el 97,6% de ellas permite organizar las búsquedas en función de las distintas categorías de los metadatos, solo habiendo un repositorio que no lo permite.

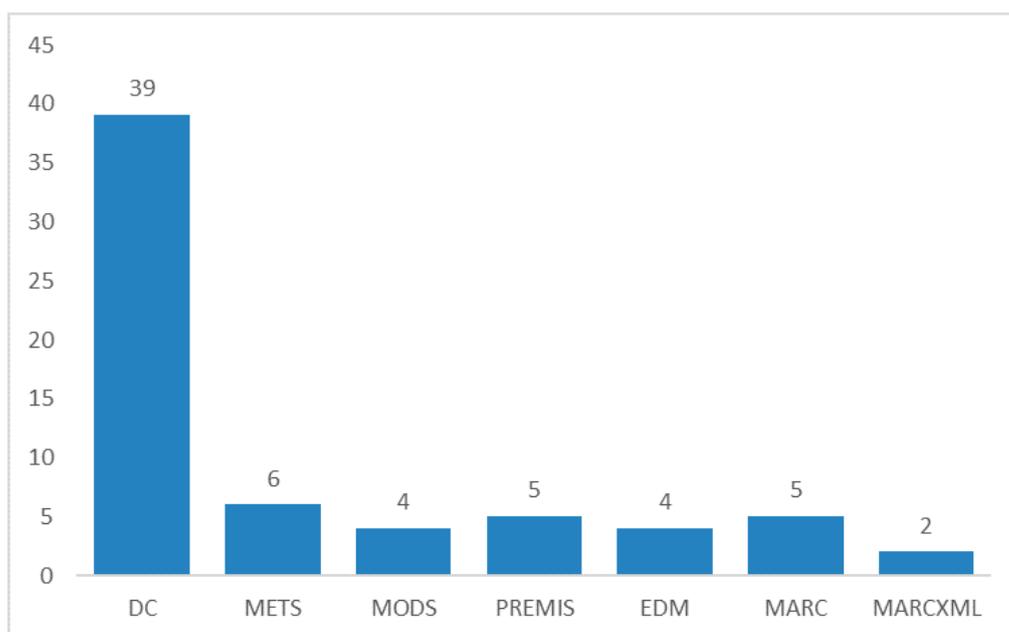
El indicador sobre accesibilidad (i09), referido a si se puede acceder o no a los documentos de los conjuntos de datos, de los treinta y nueve repositorios que alojan datos, treinta y ocho permiten el acceso a todos sus datos de investigación directamente sin ninguna restricción.

La situación de las licencias de derechos de autor asociadas a los *datasets* (i.10): treinta y siete repositorios usan alguna clase de licencia Creative Commons; y prácticamente todos indican de manera clara la licencia que usan sus datos.

La visibilidad del repositorio se conecta con la vinculación del repositorio a los diferentes recolectores (i.11): todos los repositorios con datos de investigación se vinculan al recolector RECOLECTA, treinta y siete lo hacen a Google Scholar, treinta y cuatro a BASE, treinta y dos a OpenAIRE, nueve a WorldCat y siete a Core. La mayoría de los repositorios están indexados en más de cuatro recolectores, es la situación de veintiocho de treinta y nueve repositorios que albergan datos (72%).

El último indicador analiza la existencia de guías sobre publicación de datos abiertos (i.12), veintiséis de los repositorios con datos de los treinta y nueve repositorios que albergan datos de investigación en su repositorio institucional tienen alguna guía de este tipo. Se observa que hay ocho instituciones que no tienen sección para datos ni datos de

Figura 9. Número de repositorios según tipos de metadatos usados.



investigación alojados, pero que cuentan con una guía de publicación de datos.

En general, observando las características de los repositorios institucionales que albergan datos hay una fuerte homogeneidad en la mayoría de aspectos: uso de un software vinculado al movimiento *Open Access*, plena accesibilidad a sus datos en sí, capacidad para organizar por categorías de metadatos, algo que se vincula en muchos casos al tipo de software, óptima facilidad para acceder al tipo de licencia que se usa en cuanto a los derechos de autor, así como, hay un buen resultado en la vinculación a un mínimo de cuatro recolectores, que influye en un buen nivel de interoperabilidad.

En conjunto, los resultados característicos muestran un nivel de funcionalidad bueno de los repositorios institucionales que albergan datos, pero dos aspectos se plantean en este punto para cuestionar esta situación: los conjuntos de datos de investigación, que son bajos y heterogéneos, y los tipos de metadatos. Puesto que, en cuanto a los metadatos, un 76,92% solo presentan metadatos en formato Dublin Core.

5. DISCUSIÓN

Los resultados analizados muestran que la presencia de datos de investigación en los repositorios institucionales españoles es limitada, solo cuatro universidades españolas tienen más de doscientos datos de investigación alojados en su repositorio.

Si se tiene en cuenta que un 65% de los repositorios analizados presentan políticas de acceso abierto, y que solo un 9% ha desarrollado políticas específicas para la gestión de datos de investigación, se dibuja en estos resultados una brecha entre el desarrollo de las políticas de acceso abierto y las políticas para los datos de investigación. Puede que haya que analizar otro tipo de repositorio para observar el comportamiento de los datos de investigación en el contexto español, sean estos repositorios temáticos o centralizados externos a la institución.

Sin embargo, esta falta de políticas específicas refleja una falta de desarrollo del *Data Sharing* y de la *Data Curation* por las universidades y las bibliotecas académicas.

Uno de los aspectos cruciales de la *Data Curation* es la definición de políticas y directrices para la gestión de datos de investigación. Sin políticas y prácticas establecidas, es más probable que los datos no sean accesibles, se pierdan o se gestionen de manera inadecuada. Ya en el año 2016 la IFLA inició el Data Curator Project que buscaba identificar las responsabilidades clave de los curadores de

datos y desarrollar un glosario que debería ayudar a definir mejor la profesión y desarrollar planes de estudio educativos adecuados. Pero los resultados muestran que la mayoría de las instituciones analizadas carecen de políticas específicas de gestión de datos de investigación, lo que resalta la necesidad de un mayor énfasis en la *Data Curation* en el ámbito universitario sea cual sea el tipo de repositorio en que se alojen.

Hay dos situaciones diferentes observadas en las comunidades autónomas con más datos, la Comunidad de Madrid y Cataluña.

La situación del Consorcio Madroño muestra que todas sus universidades tienen políticas de gestión de datos de investigación, siendo políticas que se vinculan al consorcio. Los datos que se vinculan a las instituciones universitarias de este consorcio se encuentran en un repositorio común que gestiona el Consorcio Madroño, pero que las propias universidades remiten a él para la observación de sus datos, siendo la opción que los repositorios institucionales permiten. Toda esta propuesta se vincularía con una práctica clara de *Data Curation* por parte de estas universidades vinculadas al consorcio. Sin embargo, esta práctica no se aplica en toda la Comunidad de Madrid, puesto que no todas las universidades forman parte del consorcio.

La situación del Consorci de Serveis Universitaris de Catalunya (CSUC) con respecto a la gestión de datos de investigación muestra que hay universidades catalanas que siguen alojando datos de investigación en sus repositorios, cuando ya se ha planteado la creación de un repositorio específico de datos. El consorcio catalán había analizado a las distintas instituciones catalanas para alinear las políticas (Alcalá y Anglada, 2019), lo que muestra un interés hacia un desarrollo de la *Data Curation*. Se plantearon dos soluciones a partir de la observación de la situación: usar algún repositorio existente, o bien, seguir la política adoptada por algunas universidades, que era la de adaptar sus repositorios institucionales para depositar datos. Finalmente, se decide crear un repositorio específico para las universidades catalanas para este cometido, y es un perfecto ejemplo para ilustrar la situación de transición en la que se encuentra la *Data Sharing* de investigación en España.

La vinculación entre la investigación, el *Data Sharing* y la *Data Curation* requiere de un enfoque más integral en la gestión de datos de investigación en las instituciones académicas, incluyendo la implementación de políticas específicas, la promoción de buenas prácticas específicas de la *Data Curation* y la adopción de estándares de metadatos más adecuados. Puesto que la *Data Curation* también

está estrechamente relacionada con la elección de estándares de metadatos, es significativo que el 76% de los repositorios con datos de investigación alojados solo usan Dublin Core.

6. CONCLUSIONES

La mayoría de los repositorios institucionales se vinculan a la ciencia abierta, considerando que presentan políticas de acceso abierto. Mientras que la mayoría no desarrolla políticas de gestión de datos vinculantes, dos tercios presentan guías informativas para la gestión de datos. Teniendo en cuenta que los datos de investigación se vinculan a la ciencia abierta misma, y que la producción científica está muy vinculada a las universidades, aunque se planteara un depósito en otro tipo de repositorios, el *Data Curation* recomienda a las universidades y a las bibliotecas académicas ser el intermediario entre el científico y el *datasharing*. Por lo que, en estos aspectos de la gestión de datos se observa un espacio posible para la mejora.

Si se atiende a los repositorios institucionales que albergan datos, en cuanto a sus características se dibuja una fuerte homogeneidad en aspectos que siguen buenas prácticas recomendadas para este tipo de plataformas. Los resultados característicos muestran un nivel de funcionalidad bueno de los repositorios institucionales que albergan datos, pero se señalan dos aspectos a mejorar: la situación de los conjuntos de datos de investigación, y los tipos de metadatos, una mayoría solo usan el formato Dublin Core, mostrando también un posible recorrido para la mejora. Actualmente, en algunos repositorios institucionales hay datos de investigación abiertos, pero es un número muy bajo.

Los conjuntos de datos se presentan en estos repositorios institucionales que albergan datos como pocos cuantitativamente en comparación con las publicaciones científicas, así como, con los datos de investigación que deberían estar asociados a la investigación española. Además, se muestra una pobre perspectiva de continuidad, si atendemos a las políticas de gestión de datos. Seis de los siete que tienen esta política se vinculan a un consorcio, que es desde donde se desarrollan estas buenas prácticas.

La situación en 2022 de los datos de investigación abiertos en repositorios de las universidades españolas sugiere que la situación es de transición, como a la espera de unas políticas aplicadas y prácticas que planteen en nuestro contexto dónde y cómo se van a almacenar los datos de investigación, y concretamente, qué va a pasar con los datos de investigación que actualmente están en algunos repositorios institucionales.

7. CONTRIBUCIÓN A LA AUTORÍA

Pablo Monteagudo: conceptualización, investigación, metodología, supervisión, validación, visualización, redacción-borrador original, redacción-revisión y edición.

Juan José Prieto: investigación, supervisión, validación, visualización, redacción-borrador original, redacción-revisión y edición.

8. DECLARACIÓN DE CONFLICTO DE INTERESES

Los autores de este artículo declara no tener conflictos de intereses financieros, profesionales o personales que pudieran haber influido de manera inapropiada en este trabajo.

9. REFERENCIAS

- Abadal, E., Ollé Castellà, C., Abad-García, F., y Melero, R. (2013). Políticas de acceso abierto a la ciencia en las universidades españolas. *Revista Española de Documentación Científica*, 36(2), e007. DOI: <http://dx.doi.org/10.3989/redc.2013.2.933>
- Alcalá, M., y Anglada, L. (2019). FAIR x FAIR. Una aproximación pragmática a la gestión de datos de investigación en modo FAIR. *Anuario ThinkEPI*, 13. DOI: <https://doi.org/10.3145/thinkepi.2019.e13e05>
- Barrueco, J. M., Rico-Castro, P., y Bonora Eve, L. V. (2021). *Guía para la evaluación de Repositorios institucionales de Investigación*. Disponible en: <https://repositoriorebiun.org/handle/20.500.11967/809>
- Clabo, N., y Ramos-Vielba, I. (2015). Reutilización de datos abiertos en la administración pública en España y uso de licencias-tipo. *Revista Española de Documentación Científica*, 38(3), e097-e097. DOI: <https://doi.org/10.3989/redc.2015.3.1206>
- d'Aquin, M., Kirstein, F., Oliveira, D., Schimmler, S., y Urbanek, S. (2023). FAIREST: A Framework for Assessing Research Repositories. *Data Intelligence*, 5(1), 202-241. DOI: https://doi.org/10.1162/dint_a_00159
- Gobierno de España. (2007). *Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público*. BOE, 276, 17 de noviembre de 2007. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2007-19814> [Fecha de acceso: 10/11/2023]
- Gobierno de España. (2011). *Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación*. BOE, 131, de 2 de junio de 2011. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2011-9617> [Fecha de acceso: 10/11/2023]
- Gobierno de España. (2015). *Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público*. BOE, 164, de 10 de julio de 2015. Disponible en: <https://www.boe.es/buscar/doc.php?id=BOE-A-2015-7731> [Fecha de acceso: 10/11/2023]

- Gobierno de España. (2022). *Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. BOE, 214, de 6 de septiembre de 2022*. Disponible en: <https://www.boe.es/eli/es/l/2022/09/05/17/> con [Fecha de acceso:10/11/2023]
- Gobierno de España. (2023). Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario. BOE, 70, de 23 de marzo de 2023. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2023-7500> [Fecha de acceso:10/11/2023]
- González-Teruel, A., López-Borrull, A., Santos-Hermosa, G., Abad-García, F., Ollé, C., y Serrano-Vicente, R. (2022). Drivers and barriers in the transition to open science: the perspective of stakeholders in the Spanish scientific community. *Profesional de la información*, 31(3), e310305. DOI: <https://doi.org/10.3145/epi.2022.may.05>
- Guallar, J; Leiva-Aguilera, J. (2013). *El content curator. Guía básica para el nuevo profesional de internet*. Barcelona: UOC.
- Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51(7-8), 662-672. DOI: <https://doi.org/10.1080/01930826.2011.601269>
- Hernández-Pérez, T. (2016). En la era de la web de los datos: Primero datos abiertos, después datos masivos. *Profesional de la información*, 25(4), 517-525. DOI: <https://doi.org/10.3145/epi.2016.jul.01>
- Hernández-Pérez, T., y García-Moreno, M.A. (2013). Datos abiertos y repositorios de datos: Nuevo reto para los bibliotecarios. *Profesional de la información*, 22(3), DOI: <https://doi.org/10.3145/epi.2013.may.10>
- Howard Hughes Medical Institute. (2003). *Bethesda Statement on Open Access Publishing*.
- IFLA (2016). *Data Curator Project*. Disponible en: https://www.ifla.org/wp-content/uploads/2019/05/assets/library-theory-and-research/Projects/firstreportprojectdatacurator.odt_.pdf
- Lyon, L. (2016). Transparency: the emerging third dimension of open science and open data. *Liber quarterly*, 25(4). DOI: <http://dx.doi.org/10.18352/lq.10113>
- Martín González, M. Y., y Ríos Hilario, A. B. (2018). *Datos académicos abiertos en las universidades españolas: Identificación, análisis y buenas prácticas*. Ediciones Universidad de Salamanca (España). Disponible en: <https://gedos.usal.es/handle/10366/138586>
- Max Planck Gesellschaft. (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*.
- Open Society Institute, (2002). *Budapest Open Access Initiative*.
- Sixto-Costoya, A., Aleixandre-Benavent, R., Vidal-Infer, A., Lucas-Dominguez, R., y Castelló-Cogollos, L. (2019). *Data sharing: Qué son y cómo se pueden compartir los datos de investigación. Manual de recomendaciones para gestores de la información*. Sociedad Española de Documentación e Información Científica.
- Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., López, A., y Rehwald, S. (2020). When are researchers willing to share their data? Impacts of values and uncertainty on open data in academia. *PLoS one*, 15(7), e0234172. DOI: <https://doi.org/10.1371/journal.pone.0234172>
- Tammaro, A. M., Matusiak, K. K., Sposito, F. A., y Casarosa, V. (2019). Data curator's roles and responsibilities: An international perspective. *Libri*, 69(2), 89-104. DOI: <https://doi.org/10.1515/libri-2018-0090>
- Torres-Salinas, D. (2010). Compartir datos (data sharing) en ciencia: contexto de una oportunidad. *Anuario ThinkEPI*, 258-261
- Unión Europea. (2019). *Directiva 2019/1024 del Parlamento Europeo y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público. Diario Oficial de la Unión Europea L172/56, de 26 de junio de 2019*. Disponible en: <https://www.boe.es/doue/2019/172/L00056-00079.pdf> [Fecha de acceso: 10/11/2023]