

Estudio de caso sobre los artículos publicados en acceso abierto con datos asociados en *PubMed Central* en un hospital universitario (2020-2024)

Blanca San-José-Montano

Hospital Clínico San Carlos
Biblioteca de Ciencias de la Salud (Madrid)
<https://orcid.org/0000-0002-5365-1590>
<https://www.directorioexit.info/ficha1951>
blanca.sanjose@salud.madrid.org

Rafael García-Carretero

Hospital Universitario de Móstoles (Madrid)
<https://orcid.org/0000-0001-7532-4585>
rafael.garcia.carretero@urjc.es

Manuel Espantaleón-Agreda

Hospital Clínico San Carlos
Biblioteca de Ciencias de la Salud (Madrid)
<https://orcid.org/0000-0002-0143-9731>
<https://www.directorioexit.info/ficha2317>
manuel.espantaleon@salud.madrid.org

San-José-Montano, Blanca; García-Carretero, Rafael; Espantaleón-Agreda, Manuel (2026). "Estudio de caso sobre los artículos publicados en acceso abierto con datos asociados en PubMed Central en un hospital universitario (2020-2024)". En: Orduña-Malea, Enrique; Olea, Isabel (coords). *El valor de las revistas científicas* (pp. 186-199). Ediciones Profesionales de la Información.

<https://doi.org/10.3145/vrc.25>

Resumen

La publicación de artículos y datos científicos en acceso abierto promueve la transparencia, la reutilización y la colaboración en el ámbito científico, destacando el papel del repositorio como *PubMed Central* (PMC) en las ciencias biomédicas. Desde 2018, PMC ha incorporado el filtro “associated data” y estrategias específicas para identificar artículos que incluyen materiales suplementarios, declaraciones de disponibilidad de datos o citaciones formales de conjuntos de datos. Este estudio tiene como objetivo analizar cómo se registran los datos asociados en los artículos de acceso abierto publicados en PMC por autores del *Hospital Clínico San Carlos* durante los últimos cinco años, así como caracterizar la tipología de datos, su declaración y su registro en repositorios. Para ello, se realizaron búsquedas filtradas en PMC, seguidas de un análisis cuantitativo y una revisión manual del texto completo de los artículos con declaración de datos, empleando como referencia las recomendaciones editoriales de MDPI. Los resultados evidencian un compromiso creciente de los autores con las políticas de ciencia abierta, aunque persisten limitaciones en cuanto a la disponibilidad, citación formal y depósito de datos en repositorios de acceso abierto. Si bien cerca del 50% de los artículos presentan datos asociados o declaraciones de datos, una proporción significativa no los registra en plataformas accesibles ni garantiza su trazabilidad. Estas deficiencias ponen de manifiesto la necesidad de avanzar en la estandarización editorial, así como de fortalecer el apoyo institucional para promover una cultura investigadora más abierta, transparente y reproducible.

Palabras clave

Ciencia abierta; Publicación acceso abierto; Repositorios; Bibliotecas de ciencias de la salud; *PubMed Central*; Filtros de búsqueda; Datos asociados; Declaración de datos.

Abstract

The publication of scientific articles and data in open access enhances transparency, reuse, and collaboration within the scientific community, highlighting the role of repositories such as *PubMed Central* (PMC) in the biomedical sciences. Since 2018, PMC has implemented the “associated data” filter, along with specific strategies to identify articles that include supplementary materials, data availability statements, or formal dataset citations. This study aims to analyse how associated data are recorded in open access articles published in PMC by authors from *Hospital Clínico San Carlos* over the past five years, as well as to characterize the types of data, their declarations, and their registration in repositories. Filtered searches were conducted in PMC, followed by quantitative analysis and manual review of the full text of articles containing data availability statements, using MDPI’s editorial guidelines as a reference. The results reveal a growing commitment among authors to open science policies; however, notable limitations remain regarding the availability, formal citation, and deposition of data in open access repositories. Although approximately 50% of the articles include associated data or data statements, a significant proportion do not register these data on accessible platforms or ensure their traceability. These shortcomings underscore the need to further standardize editorial practices and strengthen institutional support to foster a more open, transparent, and reproducible research culture.

Keywords

Open science; Open access-publishing; Repositories; Health sciences libraries; *PubMed Central*; Search filter; Associated data; Data availability statement.

1. Introducción

En el actual contexto de ciencia abierta, dos de sus pilares más importantes son la publicación de trabajos científicos en acceso abierto y la disponibilidad de los datos de investigación utilizados en su elaboración. Estos elementos buscan promover la accesibilidad, la reutilización y la colaboración en la investigación científica, en beneficio tanto de la ciencia como de la sociedad. Uno de los repositorios que más impulsa la ciencia abierta es *PubMed Central (PMC)*, especialmente en el ámbito biomédico y de las ciencias de la vida. *PMC* no sólo facilita el acceso libre al texto completo de artículos científicos —promoviendo su difusión, particularmente a través de la base de datos *PubMed*—, sino que también garantiza la preservación digital a largo plazo. El material incluido en *PMC* abarca trabajos financiados con fondos públicos de Estados Unidos, artículos publicados en acceso abierto, así como revistas completas que han establecido acuerdos editoriales con la *National Library of Medicine (PMC)* para depositar automáticamente todos o parte de sus contenidos. Estas revistas deben cumplir con ciertos criterios editoriales, contar con políticas de revisión por pares y garantizar el acceso abierto, entre otros requisitos.

En 2018 la *NLM* creó nuevos filtros para facilitar la localización de artículos publicados con material asociado en *PMC* y en *PubMed*. (*NLM*, 2018) Este nuevo filtro denominado “associated Data” permitía limitar los resultados de las búsquedas a artículos que tenían enlaces de datos relacionados con el campo *Secondary Source ID (SI)* o el *LinkOut* (más recursos). En el campo *SI* se puede localizar información/datos de los identificadores alfanuméricos de *Medline Databank Sources* que se han utilizado en la creación del artículo, como son: datos de secuencias moleculares, datos de expresión genética/abundancia molecular, números de ensayos clínicos, identificadores de *PubChem*, identificadores de *BioProject* (en la lista no figura ninguna organización ni institución española). También se localizan identificadores de los dataset (conjunto de datos) de los artículos incorporados a los repositorios de datos científicos como son repositorio *Digital Dryad* y de *Figshare* (a partir del verano de 2014). Y en el campo de *LinkOut* se pueden localizar enlaces a proveedores de texto completo, bibliotecas y bases de datos, conjuntos de datos, etc.

El uso de este filtro es diferente en ambas plataformas. En *PubMed* se puede limitar la búsqueda a trabajos que contengan materiales asociados bien en el campo *SI* como en *Linkout* (es decir no puedes limitar sólo a datos asociados formados por datasets). En *PMC* (repositorio digital de artículos de revistas a texto completo) se puede hacer la búsqueda también limitándose a localizar artículos con datos asociados en general (“*has associated data[filter]*”) o especificando mucho más el tipo de material asociado. Permite localizar artículos con material suplementario (“*has suppdata[filter]*”), artículos que tienen los datos de su elaboración disponibles o una declaración de datos (forma de obtención, por qué no pueden facilitarse, etc.) (“*has data avail[filter]*”) o trabajos con la citación bibliográfica de los datos utilizados en su elaboración “*has data citations[filter]*”.

El *Hospital Clínico San Carlos* es un hospital centro de referencia en la formación de profesionales sanitarios de la *Universidad Complutense de Madrid*. Cuenta con una amplia trayectoria de actividad investigadora liderada por el *Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC)*. Adaptándose a los nuevos mandatos nacionales y europeos (Ley 17/2022, de la Ciencia, la Tecnología y la Innovación (España, 2022); Estrategia Nacional de Ciencia Abierta 2023-2027 (MCIU, s.f.a); Ley Orgánica 2/2023, del Sistema Universitario (España, 2023); Plan Estatal de Investigación Científica y Técnica y de Innovación (PEICTI) (MCIU, s.f.b))... Los profesionales del *Hospital* publican sus trabajos en acceso abierto de publicación y datos. Con el fin de tener una descripción de cómo se está gestionando la publicación del conjunto de datos estructurados (datasets) de los artículos, tanto por los autores como por las revistas, editoriales y bases de datos, nos planteamos el siguiente trabajo.

El objetivo principal de este trabajo es conocer cómo registra los artículos *PMC* cuando establece que tienen datos asociados a través de analizar los artículos publicados en acceso abierto incorporados a esta base de datos por autores del *Hospital Clínico San Carlos* durante los últimos 5 años. Y como objetivo secundario, conocer la tipología de esos datos asociados, su tipo de declaración y si han sido registrados en algún repositorio (tipo de repositorio y cita) a través de la minuciosa revisión de cada uno de sus registros.

2. Método

Se realiza una búsqueda en 2024 y se repiten en abril 2025 (actualización de datos) en *PMC* para estudiar los resultados de los artículos publicados en este repositorio por profesionales del *Hospital Clínico San Carlos* en acceso abierto y que incluyen los datos asociados. Para ello, utilizamos la búsqueda avanzada, limitando en el campo de la “Affiliation” por “*Hospital Clínico San Carlos*” (como recomienda la firma institucional el *IdISSC* (s.f.)). A la búsqueda se le añaden los filtros “*article attributes*” como son “*open access*” y “*associated data*” durante los últimos 5 años (enero 2020 a diciembre 2024). Y, una vez establecida la estrategia de búsqueda, incorporamos diferentes filtros para obtener resultados para datos asociados en general (*has associated data[filter]*), con datos disponibles o declaración de datos (*has suppdata[filter]*) y con citación de datos (*has data citations[filter]*).

Los resultados totales de artículos científicos publicados con datos asociados es de 989, eliminamos 24 artículos (3 trabajo que pertenecían a 2019, 17 a 2025 y 4 que eran preprints de *bioRxiv* y *medRxiv*) quedando un total de 965 artículos, de los cuales iniciamos un análisis cuantitativo de su distribución por años y también se identificó a las revistas donde se han publicado artículos con mayor frecuencia (más de nueve veces) y se buscó información sobre *International Standard Serial Number (ISSN)* de papel o electrónico, editorial, tipo de licencia de publicación y si los artículos publicados analizados tenían declaración de datos.

Después pasamos a revisar manualmente el texto completo todos los artículos que llevaban datos asociados con declaración de datos (958 se eliminaron 4 que no llevaban). En cada artículo se revisaba las secciones de “material suplementario” para corroborar que llevaba material y la de “datos asociados” para comprobar su declaración de datos. Para homogeneizar la recogida de esta información, tomamos como ficha de recogida las recomendaciones de la política editorial de *MDPI* a sus autores sobre cómo establecer la declaración de disponibilidad de datos en sus manuscritos (*MDPI*, s.f.). Estas recomendaciones establecen 9 formas de declaración, que pueden verse en la tabla 1, las cuales consideramos son bastante exhaustivas y recopilan una variada tipología de declaración que engloba muy bien todos los casos que nos fuimos encontrando. Y todos los trabajos que localizamos que tenían datos en un repositorio de dominio público, los revisamos y añadimos si llevaba el enlace/nº registro, la URL, un handle, DOI o cita bibliográfica completa.

Tabla 1. *MDPI* traducción al castellano, de sus recomendaciones para autores sobre cómo declarar los datos de los manuscritos enviados para su revisión

Estado de disponibilidad de los datos	Declaración de disponibilidad de datos recomendada
Datos disponibles en un repositorio de acceso público	Los datos originales presentados en el estudio están disponibles abiertamente en [nombre del repositorio, por ejemplo, FigShare] en [DOI/URL] o [número de referencia/acceso].
Datos disponibles a pedido debido a restricciones (por ejemplo, privacidad, razones legales o éticas)	Los datos presentados en este estudio están disponibles a solicitud del autor correspondiente debido a (especifique el motivo de la restricción).
Datos de terceros	La disponibilidad de estos datos está sujeta a restricciones. Los datos se obtuvieron de [tercero] y están disponibles [de los autores/en la URL] con la autorización de [tercero].

Estado de disponibilidad de los datos	Declaración de disponibilidad de datos recomendada
Embargo de datos debido a restricciones comerciales	Los datos que respaldan los hallazgos estarán disponibles en [nombre del repositorio] en [enlace URL / DOI] luego de un embargo a partir de la fecha de publicación para permitir la comercialización de los resultados de la investigación.
Se aplican restricciones a los conjuntos de datos.	Los conjuntos de datos presentados en este artículo no están disponibles fácilmente porque [incluya el motivo, p. ej., los datos forman parte de un estudio en curso o debido a limitaciones técnicas o de tiempo]. Las solicitudes de acceso a los conjuntos de datos deben dirigirse a [entrada de texto].
Datos derivados de recursos de dominio público	Los datos presentados en este estudio están disponibles en [nombre del repositorio] en [URL/DOI], número de referencia [número de referencia]. Estos datos se obtuvieron de los siguientes recursos disponibles en el dominio público: [lista de recursos y URL].
El intercambio de datos no es aplicable (sólo es apropiado si no se generan datos nuevos o el artículo describe una investigación completamente teórica)	En este estudio no se crearon ni analizaron nuevos datos. El intercambio de datos no es aplicable a este artículo.
Los datos están contenidos en el artículo o material complementario.	Las contribuciones originales presentadas en este estudio se incluyen en el artículo/material suplementario. Para más información, contacte con el/ los autores(es) correspondiente(s).
Conjunto de datos disponible a solicitud de los autores	Los datos brutos que respaldan las conclusiones de este artículo estarán disponibles a solicitud de los autores.

3. Resultados

3.1. Búsqueda en PMC

Los resultados de las diferentes búsquedas con sus estrategias lo vemos en la tabla 2, donde también hemos añadido en la última columna los resultados finales localizados en el trabajo de campo del trabajo de campo (revisión por artículo a texto completo).

Tabla 2. Diferentes estrategias empleadas por filtro y su volumen de resultados y los resultados obtenidos de la revisión manual

Tipo de filtro	Estrategia	Resultado PMC	Resultado revisión
Publicado en acceso abierto de 2020 a 2024	"hospital clinico san carlos"[Affiliation] AND ("open access"[filter] AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate]))	2.031	---
Con datos asociados y publicado en acceso abierto, 5 años	"hospital clinico san carlos"[Affiliation] AND ("open access"[filter] AND "has associated data"[Filter]) AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate])) "hospital clinico san carlos"[Affiliation] AND "has associated data"[Filter] AND ("open access"[filter] AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate]))	989 (48,7% de 2.031)	965 (eliminado 24 trabajos)
Con datos asociados y material complementario asociado	"hospital clinico san carlos"[Affiliation] AND has suppdata[filter] AND ("open access"[filter] AND "has associated data"[Filter]) AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate]))	556 (56,2% de 989)	521
Datos disponibles o declaración de datos	"hospital clinico san carlos"[Affiliation] AND has data avail[filter] AND ("open access"[filter] AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate]))	988 (99,8% de 989)	962 (eliminado 4 sin declaración de datos) quedan 958
Citación de datos	"hospital clinico san carlos"[Affiliation] AND has data citations[filter] AND ("open access"[filter] AND "has associated data"[Filter]) AND ("2020/01/01"[PubDate] : "2024/12/31"[PubDate]))	2	1 cita datos creados; otro cita fuente datos terceros

3.2. Características trabajos publicados recuperados: material suplementario y material asociado

La distribución de trabajos publicados por año, con material suplementario, por declaración de datos y citas, podemos verlo en la figura 1, donde apreciamos que existe un aumento de su publicación en el año 2021 al 2023. El descenso del año 2024 puede estar en relación con la falta de trabajos incorporados o un retardo en dicha incorporación (a pesar de estar en 2025). De los trabajos recuperados con citas de datos asociados, sólo uno se mantiene y el otro corresponde a la cita bibliográfica de los datos con los que se había creado el trabajo, que eran fuentes del *INE* y *MCBD*.

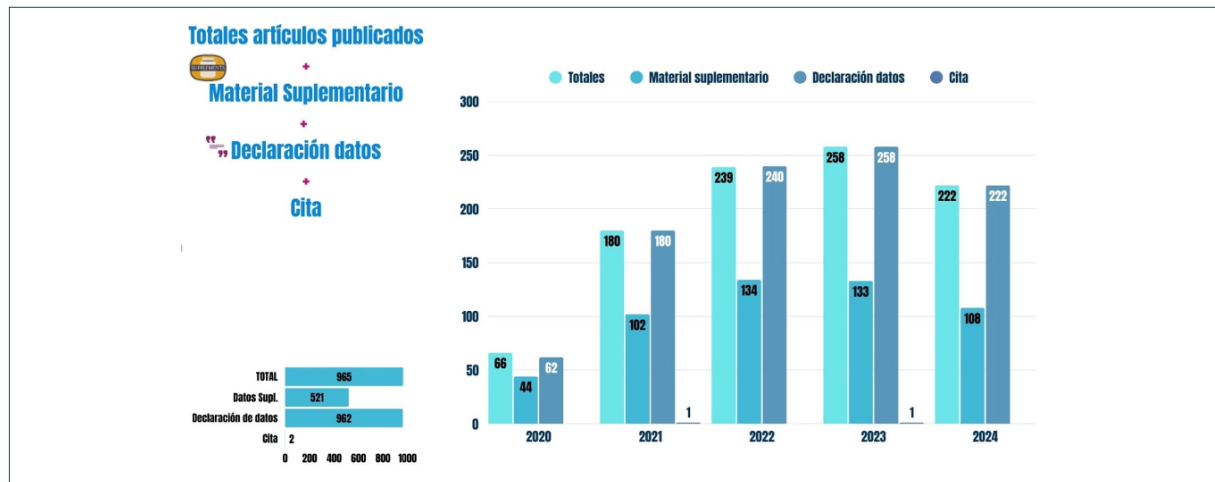


Figura 1. Distribución de trabajos publicados por años y con la diferente tipología de datos asociados

3.3. Características de las revistas donde se han publicado los trabajos

El ranking de las revistas que mayor volumen de trabajos se han publicado en acceso abierto y con datos asociados se pueden ver en la tabla 3.

Tabla 3. Ranking de revistas con mayor volumen de trabajos publicados en *PMC* (2020-2024) en acceso abierto y con datos asociados, con su ISSN, tipo de licencia, editor y declaración de datos

Ranking	Revista	ISSN	Licencia	Editor	Declaración de datos
130	<i>J Clin Med</i>	2077-0383	CC BY	MDPI	Si, 111 (85,4%)
50	<i>Int J Mol Sci</i>	1422-0067	CC BY	MDPI	Si, 42 (84%)
35	<i>Cancers</i>	2072-6694	CC BY	MDPI	Si, 31 (88,6%)
35	<i>Nutrients</i>	2072-6643	CC BY	MDPI	Si, 32 (91,4%)
35	<i>Front Immunol</i>	1664-3224	CC BY	Frontiers Media	Si, 30 (63,8%)
31	<i>Sci Rep</i>	2045-2322	CC BY	Nature Portfolio	Si, 30 (96,7%)
21	<i>Biomedicines</i>	2227-9059	CC BY	MDPI	Si, 21 (100%)
16	<i>Plos One</i>	1932-6203	CC BY	Public Library Science	Si, 16 (100%)
15	<i>Front Immunol</i>	1664-3224	CC BY	Frontiers Media	Si, 15 (100%)
13	<i>J Pers Med</i>	2075-4426	CC BY	MDPI	Si, 13 (100%)
11	<i>Front Oncol</i>	2234-943X	CC BY	Frontiers Media	Si, 10 (90,9%)
10	<i>Front Neurol</i>	1664-2295	CC BY	Frontiers Media	Si, 10 (100%)
9	<i>Cells</i>	2073-4409	CC BY	MDPI	Si, 6 (66,6%)
9	<i>Front Cardiovasc</i>	2297-055X	CC BY	Frontiers Media	Si, 9 (100%)
9	<i>Int J Environ Res Public Health</i>	1660-4601 1661-7827	CC BY	MDPI	Si, 9 (100%)

Todas las revistas publicadas en revistas de acceso abierto de vía dorada, especialmente las primeras del ranking son revistas generalistas que publican una amplia temática de trabajos. Son mayoritariamente de la editorial MDPI (8), *Frontiers Media* (5), *Public Library Science* y *Nature Portfolio* de *Springer Nature*. Todas tienen las licencias más abiertas de *Creative Commons* la CC-BY “*Creative Commons Attribution*” que permite copiar, adaptar, distribuir y usar comercialmente la obra, siempre que se reconozca la autoría. La mayoría de los trabajos tenía declaración de datos.

3.4. Título de las revistas publicadas + 9 veces distribuido a lo largo de los 5 años (2020-2024)

En las figuras 2 y 3 podemos ver la distribución por año y título de la revista con más de 4 artículos publicados al año, donde se aprecia como algunos títulos tienen al largo de todos los años un gran volumen de publicación, que va de aumento desde 2021 a 2024 como es el caso de *J Clin Med*, *Int J Molecular Sci*, *Nutrients* y *Sci Rep*. También está el caso de la revista *Int J Environ Res Public Health*, revista de MDPI que en el *Journal Citation Report* de 2023, *Clarivate* expulsó de la *Web of Science* (Nandita, 2023) y que perdió su estatus de revista con factor de impacto. De esta revista se publicaron 12 artículos en 2022 y 2023, pero en 2024 ya no se publicó ninguna.

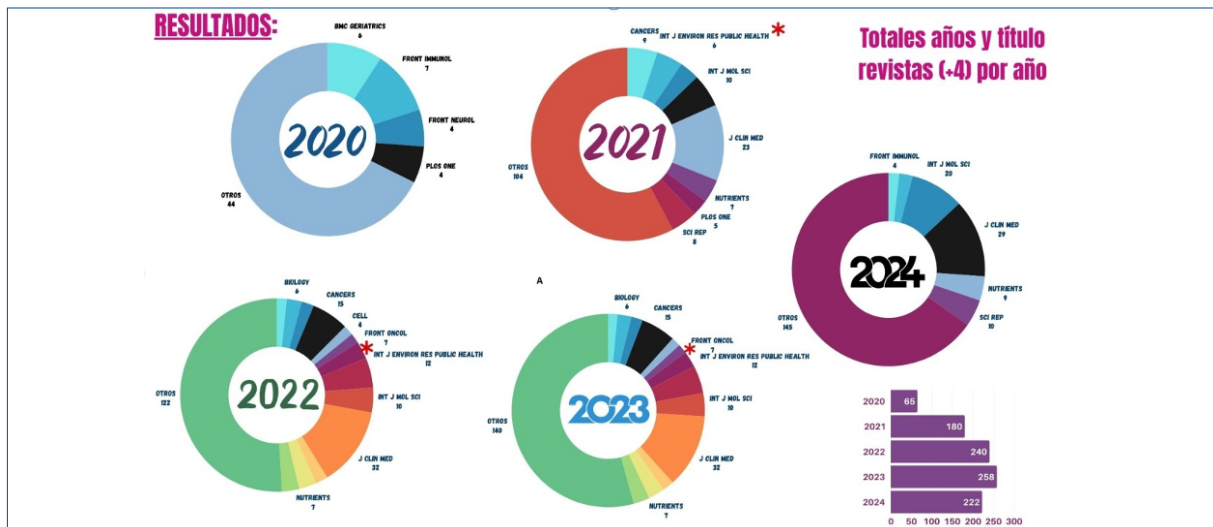


Figura 2. Distribución por año y por título de revistas en las cuales se ha publicado más de 4 artículos por años. Además, de los totales por años de las publicaciones

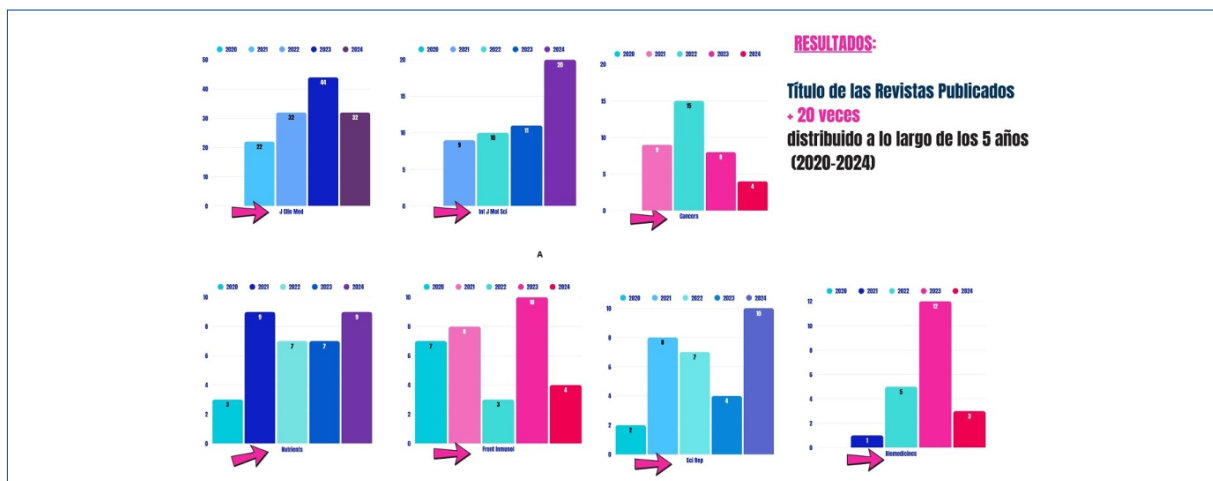


Figura 3. Título de revista publicado más de 20 veces en los últimos 5 años, distribución por año

3.5. Característica de los trabajos publicados con datos disponibles o declaración de datos

De los 962 trabajos publicados con Datos disponibles o Declaración de datos (hemos eliminado 4 trabajos por su Declaración de datos o estaba vacía o constaba como NA (*non-available*), quedando 958 los cuales llevan material suplementario 521 (el 54,4%).

Tabla 4. Tipología de la declaración de datos de los artículos revisados en el trabajo de campo (958), con sus años y su volumen de datos suplementarios

Tipo de declaración de datos	Volumen/Año		+ Datos Suplementarios
Conjunto de datos disponible a pedido de los autores	552		296
	2020	32	
	2021	104	
	2022	122	
	2023	158	
	2024	136	
Datos disponibles en un repositorio de acceso publico	62		47
	2020	7	
	2021	15	
	2022	20	
	2023	9	
	2024	11	
Datos derivados de recursos de dominio público	8		4
	2020	1	
	2021	4	
	2022	1	
	2023	1	
	2024	1	
Datos disponibles a pedido debido a restricciones (por ejemplo, privacidad, razones legales o éticas)	56		32
	2020	7	
	2021	9	
	2022	15	
	2023	12	
	2024	13	
El intercambio de datos no es aplicable	117		38
	2020	3	
	2021	17	
	2022	33	
	2023	39	
	2024	25	
Los datos están contenidos en el artículo o material complementario	129		81
	2020	10	
	2021	20	
	2022	30	
	2023	35	
	2024	34	
Se aplican restricciones a los conjuntos de datos	34		23
	2020	3	
	2021	8	
	2022	13	
	2023	8	
	2024	2	
	958		521

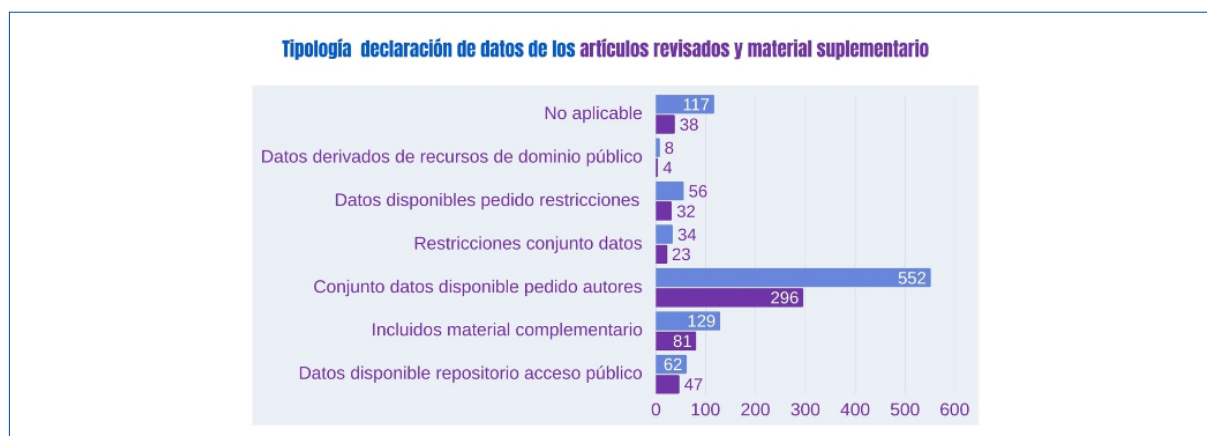


Figura 4. Tipología de declaración de datos de los artículos revisados y material suplementario

- En el 12,2% (117) el intercambio de datos no era aplicable por su tipología de trabajo (caso clínico, revisión, etc.),
- Solo un 0,8% (8) utilizaban datos derivados de recursos de dominio público, como los procedentes del *Instituto Nacional de Estadística (INE)* y del *Sistema de Información Sanitaria del Ministerio de Sanidad*, concretamente del *Registro de Actividad de Atención Especializada (RAE-CMBD)* y su *Conjunto Mínimo Básico de Datos de las Altas Hospitalarias*. Además, en este apartado hemos incluido aquellos artículos que declaran haber utilizado bases de datos como *PubMed* o *ClinicalTrials.gov* para realizar revisiones sistemáticas o narrativas.
- En el 5,8% (56) los datos están disponibles a pedido debido a restricciones (por ejemplo, privacidad, razones legales o éticas). Aquí hemos incorporado los que avisan de esas restricciones o la firma de un acuerdo, pero si facilitan los datos.
- El 3,5% (34) aplica restricciones al conjunto de datos hemos incorporado en esta sección cuando establecen que deben cumplir las políticas de compartir datos de un laboratorio (p.ej. *AstraZeneca*), el embargo de datos, o acuerdos de acceso especiales como el grupo *Predimed* (grupo multicéntrico de investigadores españoles “Prevención con Dieta Mediterránea”), o cuando declaran que los datos están en las bases de datos de los hospitales o notas clínicas de hospitales (p. ej. *Hospital Miguel Servet* o notas clínicas *Hospital Clínico San Carlos*).
- El mayor volumen de trabajos, un 57,6% (552) el Conjunto de datos está disponible a pedido de los autores, sin informar de cómo están esos datos, si dispone de un Plan de Gestión de datos, formatos, etc. A veces también señala el nombre del grupo de trabajo, laboratorio, instituto o fundación de investigación (p.ej. *Pethema Foundation*, *Menarini*, etc. a quien deben solicitar los datos.
- Un 13,5% (129) los datos están contenidos en el artículo o material complementario
- Y por último, hemos localizado un 6,57% (62) que los datos están disponibles en un repositorio de acceso público. En este apartado hemos incluido aquellos trabajos que habían depositado material de bases de datos o plataformas específicas de su especialidad, además de repositorios temáticos o institucionales.

3.6. Declaración de datos: datos disponibles en un repositorio de acceso público

En la tabla 5 se puede ver la tipología de repositorio, base de datos, portales, etc. donde se han depositado datos, los resultados muestran:

- Tan sólo dos trabajos llevan incorporado la cita completa del depósito de datos en un Repositorio temático especializado y se le ha asignado un identificador persistente (DOI). Uno de ellos recuperado en la búsqueda inicial en *PMC* y el otro descubierto en el trabajo de campo. Con los datos incorporados a Repositorios temáticos (*Dryad* y *Zenodo* que les han facilitado DOI).
- Un gran volumen de trabajos incorpora sus datos muy especializados en bases de datos y archivos temáticos donde pueden compartir sus datos y generan un registro identificador, como son 165r RNA, EVA, GEO, GENBANK, EGA, etc. Además, de portales o programas donde se les facilita la gestión especial de sus datos por ejemplo OA-Breast Database o cBioportal.
- Tan sólo catorce trabajos han incorporado sus datos a un repositorio:
 - Seis *Digital CSIC*
 - Cinco en *Zenodo*
 - Uno en *Figshare*
 - Uno en *Dryad*
 - Uno en *RIMAsalud*, nuestro repositorio institucional.
- Todos, menos el que informa que los depositará de *Digital CSIC* una vez publicado, disponen de un identificador handle o en el caso de *Zenodo* de “Record” pero siendo repositorios que facilitan un DOI, no lo incorporan al trabajo publicado.
- Tan solo un trabajo ha depositado sus datos en *RIMAsalud*, nuestro repositorio Institucional.

Tabla 5. Nombre y tipología de los repositorios, bases de datos y plataformas donde los artículos analizados han depositado o registrado sus datos, indicando si facilita el acceso y la existencia citas asociadas.

Vol.	Denominación	Tipo plataforma	Acceso	Identificador persistente	Cita
9	<i>Gene Expression Omnibus (GEO)</i> ¹	Base de datos	Si, los 9	No	No
6	<i>DIGITAL CSIC</i>	Repositorio Institucional Multidisciplinar	Si, solo 6	5 con URI 2 con DOI 1 después de la publicación	No
5	<i>EMBL-EBI y EGA (European Genome-Phenome)</i> ²	Plataforma Big Data	Si, solo 3	No	No
5	<i>GeneBank</i> ³	Base de datos pública	Si, los 5	No	No
5	<i>SCOPE Proyect</i> ⁴	Proyecto	Si, los 5	No	No
5	<i>Zenodo</i>	Repositorio	Si, los 5	5 con URI	Si, solo 1
3	<i>Bioproject</i> ⁵	Colección datos	Si, 1	No	No
3	<i>CBioportal</i> ⁶	Plataforma abierta	No	No	No
2	<i>ENA</i> ⁷	Colección de biodatos	Si, 2	No	No
2	<i>Figshare</i> ⁸	Repositorio digital	Si, 2	2 con DOI	No
2	<i>OSF (Open Science Framework)</i> ⁹	Repositorio digital	Si, 2	No	No
1	<i>16S rRNA Database</i> ¹⁰	Repositorio medicina	Si, 1	Con URI y DOI	Si
1	<i>DRYAD</i>	Repositorio biomedicina	Si, 1	Con URI y DOI	Si
1	<i>European Genome-phenome Archive (EGA)</i> ¹¹	Repositorio europeo	Si, 1	No	No
1	<i>EVA (European Variation Archive)</i> ¹²	Base de datos	Si, 1	No	No
1	<i>Fundación ACE</i> ¹³	Fundación	Si, 1	No	No

Vol.	Denominación	Tipo plataforma	Acceso	Identificador persistente	Cita
1	GEPIA2 y Timer.2 ¹⁴	Herramienta web	1, si	No	No
1	GitHub ¹⁵	Plataforma	1, si	No	No
1	Hope Project ¹⁶	Organización global	1, si	No	No
1	LOVD v.3.0—Leiden Open Variation Database ¹⁷	Base de datos ADN	1, si	No	No
1	National Center for Research Resources (NCRR) ¹⁸	Instituto Nacional de Salud	1, si	No	No
1	OA-Breast Database ¹⁹	Conjunto de datos	No	No	No
1	PRIDE ²⁰	Repositorio de socios	1, si	No	No
1	RIBRON ²¹	Registro	No	No	No
1	RIMAsalud ²²	Repositorio Institucional	1, si	No	No
1	No se encuentra en el material suplementario ²³		Si, 1	No	No

Notas:

1. Base de datos específica para secuencias ARNr 16S de *National Center for Biotechnology Information (NCBI)*.
2. *European Bioinformatics Institute. Laboratorio Europeo de Biología Molecular*. Plataforma de Big Data base de datos de secuencias de nucleótidos del *EMBL*.
3. Base de datos pública de secuencias de ácidos nucleicos (ADN y ARN) mantenida por el *NCBI*.
4. *Scope* tiene como objetivo fomentar la innovación y el crecimiento empresarial en Europa mediante el desarrollo de habilidades de intraemprendimiento.
5. Colección organizada de datos biológicos relacionados con un proyecto de investigación específico, gestionada por *NCBI*.
6. Plataforma de código abierto diseñada para explorar, visualizar y analizar datos genómicos del cáncer, funciona parte como un repositorio.
7. *European Nucleotide Archive*, es una colección global de recursos de biodatos.
8. Repositorio digital para datos, artículos, presentaciones, videos, figuras, y otros tipos de contenido académico. Asigna DOI.
9. Repositorio digital por el *Center for Open Science (COS)*, datos sin procesar accesible mediante acuerdo.
10. Base de datos específica para secuencias ARNr 16S de *National Center for Biotechnology Information (NCBI)*.
11. Repositorio europeo seguro que almacena y distribuye datos genómicos y fenotípicos humanos generados por estudios biomédicos.
12. Base de datos de acceso abierto con todo tipo de datos de variación genética de todas las especies.
13. *Fundación ACE* es una entidad sin ánimo de lucro al servicio de las personas con enfermedad de Alzheimer u otras demencias.
14. (*Gene Expression Profiling Interactive Analysis v2*) Herramienta webs centrada en el análisis del microambiente inmune tumoral.
15. Plataforma alojar proyectos de software usando *Git*, un sistema de control de versiones.
16. Organización global que ofrece atención médica y ayuda humanitaria a comunidades en todo el mundo, con especial énfasis en situaciones de emergencia y crisis.
17. Base de datos que facilita la recopilación y visualización en línea de variantes de ADN centradas en los genes.
18. Institutos Nacionales de Salud (NIH) que se centró en proporcionar recursos e instalaciones de investigación compartidos para apoyar la investigación biomédica Se disolvió en 2011.
19. Conjunto de datos desarrollado por el *Computational Imaging Science Laboratory* de la *Universidad de Illinois*.
20. Consorcio *ProteomeXchange* a través del repositorio de socios, no en acceso abierto.
21. Registro Español Informatizado de Pacientes con Bronquiectasias.
22. Repositorio Institucional de la Consejería de Sanidad de la Comunidad de Madrid.
23. Los conjuntos de datos presentados en este estudio se encuentran en repositorios en línea. Los nombres de los repositorios y sus números de acceso se encuentran en el artículo / Material complementario.

4. Discusión

El análisis de las publicaciones científicas realizadas por los profesionales del *Hospital Clínico San Carlos* en *PubMed Central* revela patrones significativos en la gestión y declaración de datos asociados, especialmente en lo que respecta al acceso abierto y la disponibilidad de datos tanto por parte de los autores, las editoriales y *PMC*.

4.1. Datos disponibles, declaración de datos y bases de datos

La metodología empleada evidencia la dificultad de filtrar o localizar trabajos de investigación que presenten datos asociados, declaraciones de disponibilidad de datos o citaciones de conjuntos de datos en bases de datos. Sólo fue posible aplicar estos filtros en el repositorio *PMC*. Sin embargo, los resultados obtenidos deben ser revisados con cautela, ya que presentan inexactitudes. Entre estas se encuentran la inclusión de trabajos que no corresponden al año especificado o que carecen de la información esperada. Asimismo, se identificaron trabajos con citación de datos que no aparecían en las búsquedas, como es el caso de un artículo depositado en el repositorio *Dryad* cuya citación no figuraba en los resultados.

Cabe destacar que el concepto de “datos asociados” no se corresponde necesariamente con el de “dataset” o conjunto de datos. Incluso entre los artículos que presentan una declaración de datos, dicha información parece responder más a un estándar de respuesta predefinido que a una descripción verificable y precisa. De hecho, se detectó una contradicción significativa: el 12 % de los trabajos con declaración de datos indican que dicha declaración no es aplicable a su investigación.

4.2. Publicaciones y disponibilidad de datos

Durante el período de estudio (2020-2024), se identificaron 2.031, de los cuales el 47,5% contenían datos asociados. Este porcentaje refleja un esfuerzo considerable por parte de los investigadores en cumplir con las políticas de datos abiertos, aunque también indica que una mayoría significativa de los trabajos no proporciona acceso a los datos subyacentes. La falta de datos asociados en el 52,5% de los artículos podría estar relacionada con la diversa tipología de trabajo y también, con diversas barreras, tales como preocupaciones éticas, limitaciones técnicas o una falta de infraestructura adecuada para la planificación, gestión y almacenamiento de grandes conjuntos de datos.

De los 962 artículos que incluyen declaraciones de datos, un porcentaje importante (57,6%) los autores declaran facilitar los datos disponible previa solicitud, lo que sugiere que, aunque los autores están dispuestos a compartir sus datos, persisten restricciones significativas en cuanto al acceso abierto total. Un 5,8% los autores ya declaran que los datos están disponibles a pedido debido a restricciones (por ejemplo, privacidad, razones legales o éticas). Estas restricciones pueden deberse a la naturaleza sensible de los datos, la necesidad de acuerdos de confidencialidad o políticas institucionales que limitan la disponibilidad pública de los mismos. En contraste, tan solo un 3,5% de los artículos indican explícitamente que se aplican restricciones al conjunto de datos.

Aunque la ciencia abierta es un ideal cada vez más extendido, estos temores reflejan barreras reales que requieren soluciones prácticas y cambios culturales en el ámbito académico.

4.3. Material suplementario y citación de datos

El análisis también reveló que una proporción considerable de los artículos con datos disponibles incluye material suplementario (48,6%), lo que sugiere que los investigadores están utilizando este recurso como un medio para proporcionar información adicional sin sobrecargar el artículo principal. Incluso de los artículos con declaración de datos, un 13,5% declara incluirlos en el Material suplementario dentro del artículo.

Sin embargo, es notable que solamente dos de los artículos analizados incluyera una citación formal de los datos (uno de ellos incluso no rescatado en la búsqueda bibliográfica), lo que representa una omisión crítica en términos de buenas prácticas científicas, especialmente en la inclusión de los datos en un repositorio. La citación de datos es fundamental para la trazabilidad y la reutilización de los datos, y su ausencia puede limitar la visibilidad y el impacto de la investigación.

4.4. Impacto de las revistas y políticas editoriales

La editorial en la que se ha publicado más trabajos es *MDPI* que facilita a los autores una lista de posibles estados de disponibilidad de datos para estandarizar este tipo de información (*MDPI*, s.f.), lo que puede facilitar a los autores mejorar sus prácticas de ciencia abierta y la transparencia de su investigación y difusión. A la vez que a las editoriales les permite seguir las *Transparency and Openness Promotion Guidelines* (TOP) (COS, 2020), muy especialmente la transparencia de los datos. Además, la mayoría de las revistas más utilizadas tienen la licencia más abierta de difusión que es CC BY, y más de un 80% de estos trabajos tienen declaración de datos.

4.5. Limitaciones del estudio

Es importante destacar que este análisis se limita a las publicaciones de un único centro hospitalario (sin ser una búsqueda exhaustiva por institución), y puede no ser representativo de las prácticas en otras instituciones. Además, los trabajos recuperados en *PMC* están sujetos a los acuerdos con las editoriales para incluir los artículos de sus revistas en el repositorio con el fin de fomentar el acceso abierto y preservar sus trabajos. Actualmente un gran volumen de editoriales tiene acuerdo para la incorporación de algunas de las revistas que publican en abierto. Por ejemplo *MDPI* incorpora 92 revistas a *PMC* de las 471 que publica (*MDPI*, s.f.) y *Frontiers Media* 60 de las 230 que publica (*PMC*, s.f.).

4.6. Declaración de datos y datos asociados

De los trabajos localizados con Declaración de datos, solamente un 6,57% (62) sus datos están disponibles en un repositorio de acceso público, hemos incluido en esta definición bases de datos especializadas, portales temáticos, repositorio todos ellos de acceso abierto, pero solo 22,6% (14) trabajos han depositado sus datos en un Repositorio Institucional o generalista. Lo que muestra una asimilación cultural del depósito de los datos utilizados en el desarrollo del trabajo para ser reutilizados o para reproducir el trabajo.

5. Conclusión

En resumen, aunque hay un esfuerzo notable por parte de los autores, editores y bases de datos para cumplir con las políticas de datos abiertos, existen áreas claras de mejora, particularmente en lo que respecta a la citación de datos, facilitación de la gestión integral de los datos (depósito repositorio) y la eliminación de barreras al acceso. El aumento de la estandarización en las políticas editoriales, un mayor apoyo institucional y los Servicios de Apoyo a la Investigación de las bibliotecas podrían facilitar un avance hacia una ciencia más abierta y reproducible.

6. Notas

1. Contiene el *Genomics data repositories* (*GEO*, *EGA*, *cBioportal*, y *TCGA*, *PRIDE* datos proteómicos cuantitativos).

2. Permite a los desarrolladores almacenar, compartir y colaborar en el desarrollo de código de forma eficiente. Todos los scripts que permiten la replicación de todos los análisis están disponibles en los archivos complementarios y el repositorio público.
3. *National Center for Research Resources (NCRR)*, y sus funciones se redistribuyeron a otras entidades de los *NIH*, principalmente al *Centro Nacional para el Avance de las Ciencias Translacionales (NCATS)*.
4. *OA-Breast Database* diseñado para facilitar estudios de simulación en imágenes médicas, especialmente en modalidades como la tomografía fotoacústica (PACT), la tomografía por ultrasonido (USCT) y la tomografía óptica difusa.

7. Referencias

- COS (2020). "New measure rates quality of research journals' policies to promote transparency and reproducibility". *Center for Open Science*, February 10.
<http://www.cos.io/about/news/new-measure-rates-quality-research-journals-policies-promote-transparency-and-reproducibility>
- 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". *Boletín oficial del estado*, n. 214, 6 septiembre.
<https://www.boe.es/eli/es/l/2022/09/05/17/con>
- España (2023). "Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario". *Boletín oficial del estado*, n. 70, 23 marzo.
<https://www.boe.es/eli/es/lo/2023/03/22/2/con>
- IdISSC (s. f.). ¿Cómo citar correctamente en las publicaciones?
<https://www.idissc.org/wp-content/uploads/2022/01/2019-03-25-Como-citar-las-publicaciones.pdf>
- MDPI (s.f.). Supplementary materials, data deposit and software source code.
<https://www.mdpi.com/journal/ijerph/instructions#suppmaterials>
- MDPI (s.f.). *Journal list*.
<https://www.mdpi.com/about/journals>
- Ministerio de Ciencia, Innovación y Universidades. *Estrategia Nacional de Ciencia Abierta 2023-2027*.
<https://www.ciencia.gob.es/Estrategias-y-Planes/Estrategias/ENCA.html>
- Ministerio de Ciencia, Innovación y Universidades. *Plan Estatal de Investigación Científica y Técnica y de Innovación (PEICTI)*.
<https://www.ciencia.gob.es/Estrategias-y-Planes/Planes-y-programas/PEICTI.html>
- NLM (2018). "Data Filters in PMC and PubMed". *NLM Tech Bull*. Mar-Apr, 421, b8.
https://www.nlm.nih.gov/pubs/techbull/ma18/brief/ma18_pmc_data_filters.html
- PMC (s.f.). *Selective deposit collection*.
<https://pmc.ncbi.nlm.nih.gov/about/selectivedeposit/>
- Quaderi, Nandita (2023). "Supporting integrity of the scholarly record: Our commitment to curation and selectivity in the Web of Science". *Clarivate. Academia & Government*, March 20.
<https://clarivate.com/academia-government/blog/supporting-integrity-of-the-scholarly-record-our-commitment-to-curation-and-selectivity-in-the-web-of-science/>