

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

OPEN DATASETS AND THEIR CONTRIBUTION TO THE SOCIAL IMPACT OF SCIENCE

Paloma G. BLÁZQUEZ¹, Almudena MANGAS-VEGA² y Francisco José GARCÍA-PENÁLVO³

¹ *Universidad de Salamanca. España*

palomagb@usal.es

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

² *Universidad de Salamanca. España*

almumvega@usal.es

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

³ *Universidad de Salamanca. España*

fgarcia@usal.es

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

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

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

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

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

1. INTRODUCCIÓN

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

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

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

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

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

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

Esta concepción ha ido integrándose en las políticas científicas a través de marcos normativos que reconocen el valor de los datos como resultados científicos en sí mismos. En España, este reconocimiento se refleja en convocatorias como el programa *Generación del Conocimiento* financiado por la Agencia Estatal de Investigación (AEI)¹, donde los Planes de Gestión de Datos (PGD) son ya un requisito obligatorio. A nivel europeo, programas como Horizon Europe² refuerzan esta perspectiva, estableciendo el acceso abierto a los datos como un principio estructural de la investigación financiada con fondos públicos. La Comisión Nacional Evaluadora de la Actividad Investigadora (CNEAI)³ exige que los conjuntos de datos estén depositados en repositorios de acceso abierto, de acuerdo con lo establecido en el artículo 37.2 de la Ley 17/2022 de la Ciencia, la Tecnología y la Innovación, para que puedan ser considerados en los procesos de evaluación.

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

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

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

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

3 Ministerio de Ciencia, Innovación y Universidades. (2024, 24 de diciembre). Resolución de 11 de diciembre de 2024, de la Secretaría General de Universidades, por la que se aprueba la convocatoria de evaluación de la actividad investigadora. *BOE* núm. 309, pp. 179840-179845. Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA). https://www.aneca.es/documents/20123/238516/BOE-A-2024-27007_convotaria+SGU.pdf/943922c6-1f98-6ac4-063c-d1b1788324eb?t=1736258650652

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

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

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

2. METODOLOGÍA

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

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

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

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

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

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

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

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

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

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

3. RESULTADOS Y DISCUSIÓN

3.1. DIMENSIÓN DESCRIPTIVA: CARACTERIZACIÓN DE LOS REPOSITARIOS INSTITUCIONALES

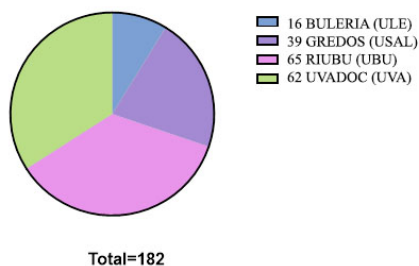
El análisis se ha centrado en los repositorios institucionales de las universidades públicas que integran el consorcio BUCLE (Bibliotecas Universitarias de Castilla y León), formado por la Universidad de Burgos (UBU), la Universidad de León (ULE), la Universidad de Salamanca (USAL) y la Universidad de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

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

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

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

Figura 1



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

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

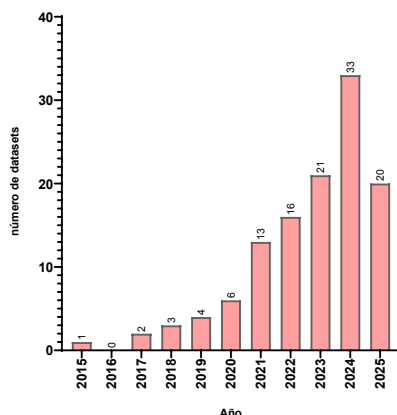
Estas carencias apuntan a una implementación todavía parcial de los principios FAIR en los repositorios institucionales. Aunque algunos requisitos formales, como el uso de DOI, se cumplen mayoritariamente, no siempre van acompañados de una documentación adecuada que permita aprovechar los datos de forma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

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

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

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

Figura 2



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

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

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

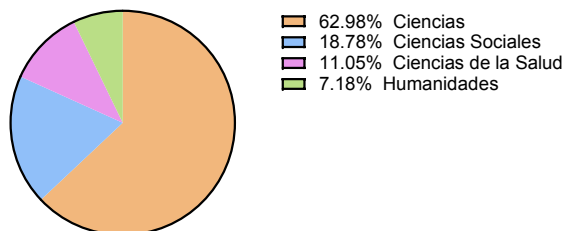
En cuanto a la frecuencia de contribución, el patrón es claramente puntual: el 75,3 % de los autores ha participado en un único depósito. Solo una minoría firma más de dos conjuntos de datos y los casos con diez o más aportaciones son excepcionales. Este comportamiento sugiere que el depósito de datos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

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

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

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

Figura 3



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

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

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

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

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

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

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

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

Tabla 1

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

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

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

3.3. DIMENSIÓN ALTMÉTRICA: COMUNICAR PARA REUTILIZAR Y TRANSFORMAR

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Otro caso relevante, esta vez fuera del ámbito sanitario, es el informe de la Oficina Nacional de Estadística del Reino Unido⁶ sobre los precios de la vivienda en pequeñas áreas de Inglaterra y Gales. Publicado en junio de 2023 y recogido en la plataforma Overton, el documento se basa de forma explícita en múltiples conjuntos de datos abiertos. En este caso, los datos no solo alimentan el análisis, sino que fundamentan decisiones en áreas como la planificación urbana, la fiscalidad o el acceso a la vivienda.

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

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

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

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

6 Office for National Statistics. (2023, 20 de septiembre). *Median house prices for administrative geographies: HPSSA dataset 9*. <https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianhousepriceforationalandsnationalgeographiesquarterlyrollingyearhpssadataset09>

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

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

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

4. CONCLUSIONES

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

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

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

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

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

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

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

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

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

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

REFERENCIAS BIBLIOGRÁFICAS

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

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

