

XXI Workshop de Proyectos Digitales REBIUN

Universidad de Castilla-La Mancha, 28 y 29 de
mayo de 2026

Del acceso al impacto: Claves para evaluar los datos de investigación

Iratxe Puebla, Make Data Count

Del acceso al impacto: Claves para evaluar los datos de investigación

Iratxe Puebla, Make Data Count Fecha

URI: <https://hdl.handle.net/20.500.11967/6745>

Progreso en políticas y prácticas de datos

Desarrollo de infraestructura para gestionar, publicar y preservar datos.

- +3000 repositorios en Re3data
- 70 millones de datos con DOIs registrados con DataCite

Desarrollo y adopción de los principios FAIR.

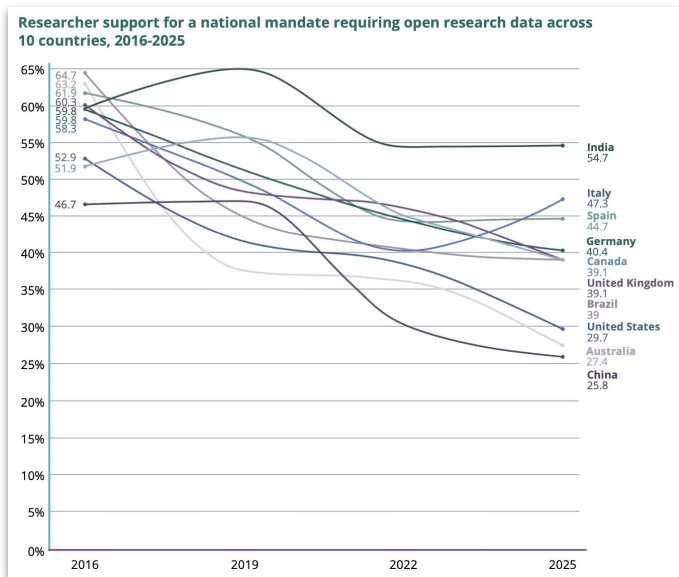
Políticas institucionales y de agencias de financiación que promueven el acceso a los datos.

Políticas en revistas y editoriales.



Aún queda camino por recorrer...

Las opiniones sobre las políticas de datos abiertos varían.



La indexación y la visibilidad amplia de los datos abiertos aún no es la norma.

→ 11,356,961 datasets in DataCite

→ 5.7 % with discipline information

→ OECD Fields of Science

→ **Natural Sciences**

Mathematics; Computer and information; Physical;

Chemical; Earth and related environmental; Biological; Other

→ **Engineering and Technology**

Civil; Electrical, electronic, information; Mechanical;

Chemical; Materials; Medical; Environmental engineering;

Environmental biotech; Industrial biotech; Nanotech; Other

→ **Medical and Health Sciences**

Basic medicine; Clinical medicine; Health sciences; Medical

biotech; Other

→ **Agricultural Sciences**

Agriculture, forestry, fisheries; Animal and dairy; Veterinary;

Agricultural biotech; Other

→ **Social Sciences**

Psychology, Economics and business; Educational; Sociology;

Law; Political; Social and economic geography; Media and

communications; Other

→ **Humanities**

History and archaeology; Languages and literature;

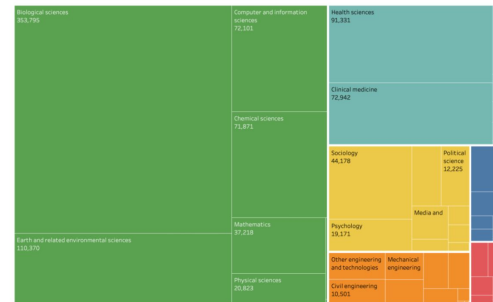
Philosophy, ethics, religion; Arts; Other

→ 0.9 % with citations

→ 0.02% with discipline and citations

Data generated on 2 April 2022 with Jupyter Notebook:

Ninkov, A. (2021). antonninkov/ISSI2021: Datasets on DataCite - an Initial Bibliometric Investigation (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5092301>



Es importante recordar las motivaciones detrás de este trabajo

Mediante datos abiertos queremos:

- Impulsar progreso en investigación e innovación
- Fomentar mayor participación en la investigación
- Aumentar la reproducibilidad de la investigación

Debemos poder determinar si estamos logrando estos objetivos.

El foco en datos abiertos se ha centrado en el **acceso**.

La asignatura pendiente:
Comprender la **utilización**
de los datos.

Comprender el uso de los datos es clave para orientar políticas, asignar recursos y dar reconocimiento a los datos



Las **agencias de financiación** y las **universidades** quieren comprender el cumplimiento de las políticas de datos y evaluar el alcance e impacto de los datos de los proyectos de investigación.



Los **repositorios** quieren maximizar la visibilidad de sus datos e informar sobre el valor y alcance de sus colecciones de datos.



L@s **investigadore/as** quieren comprender el alcance de datos y que estas contribuciones sean valoradas en la evaluación académica.



**MAKE
DATA COUNT.**

**Una iniciativa comunitaria que promueve
prácticas y herramientas para la
evaluación responsable de la utilización,
alcance e impacto de los datos.**

Make Data Count

Nuestro objetivo: un ecosistema donde los datos **se evalúan y son reconocidos** como contribuciones importantes de investigación.



Infraestructura abierta
para recopilar y compartir
medidas de utilización de
datos

Estándares y evidencia
para contextualizar las
métricas de datos

Promoción del
reconocimiento del uso e
impacto de los datos en la
evaluación académica

<https://makedatacount.org/learn-about-us/>

Infraestructura abierta para información sobre la utilización de datos



Vistas y descargas de datos

Make Data Count ha desarrollado recomendaciones para capturar y compartir vistas y descargas de los datos.

Recomendaciones basadas en el **Code of Practice for Research Data**, desarrollado por Make Data Count y COUNTER, que provee un estándar para normalizar indicadores de utilización.

- Promueve la **transparencia** en cómo se recopilan y procesan los indicadores.
- Permite la **comparación e interpretación** de diferentes grupos de datos y plataformas.

<https://makedatacount.org/explore-resources/resources-for-repositories-and-data-platforms/>

3 pasos para implementar el monitoreo del uso de los datos

Los indicadores de uso proporcionan información valiosa sobre el uso e impacto de los datos. Al implementar las recomendaciones de Make Data Count (MDC), los repositorios recopilan conteos estandarizados y avanzan el conocimiento comunitario sobre cómo se utilizan los datos.

01 Implementa las recomendaciones de MDC

Dos maneras de recopilar vistas y descargas de datos normalizadas según el estándar COUNTER:

- Usage tracker, un widget que recopila conteos en las páginas de los archivos de datos.
- Procesamiento de conteos a nivel del servidor.

02 Comparte los conteos de uso de datos

- Muestra los conteos de uso en las páginas de los datos para aumentar visibilidad y reconocimiento para los/as creadores/as de datos.
- Envía los conteos a DataCite para su agregación y descubrimiento.

03 Analiza y obtén información

- Revisa los conteos de uso para confirmar alineamiento con estándares de monitoreo o identificar valores atípicos.
- Obtén información sobre los datos más utilizados y el uso a través del repositorio a lo largo del tiempo.

Vistas y descargas de datos: Modalidades de implementación

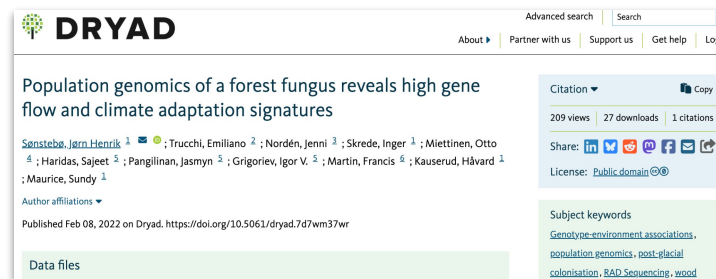
Característica	DataCite usage tracker	Log processing
Conteo de vistas y descargas	Capturados a través de las páginas de los archivos de datos	Capturado a nivel del servidor
Normalización	El usage tracker normaliza los conteos acorde al código COUNTER	El repositorio procesa los conteos acorde a las reglas del código COUNTER
Informes de utilización	Generado cada mes por el usage tracker	Generado por el repositorio o plataforma, en cualquier momento
Envío a DataCite	El usage tracker envía los informes directamente al API Usage Reports	El repositorio o plataforma completa el envío de informes

<https://makedatacount.org/explore-resources/resources-for-repositories-and-data-platforms/>

Vistas y descargas de datos

Los conteos de vistas y descargas son accesibles a través del API Usage Reports de DataCite y el portal DataCite Commons.

- Recomendamos que los repositorios también muestren las vistas y descargas en las páginas de los datos.
- Agregación mediante DataCite aumenta la visibilidad.



DRYAD Advanced search Search

About Partner with us Support us Get help Log in

Population genomics of a forest fungus reveals high gene flow and climate adaptation signatures

[Sensibar, Jörn Henrik](#) ¹ ; [Trucchi, Emiliano](#) ² ; [Nordén, Jenni](#) ³ ; [Skrede, Inger](#) ¹ ; [Miettinen, Otto](#) ⁴ ; [Haridas, Sajeet](#) ⁵ ; [Pangilinan, Jasmyn](#) ⁵ ; [Grigoriev, Igor V.](#) ⁵ ; [Martin, Francis](#) ⁶ ; [Kausarud, Håvard](#) ¹ ; [Maurice, Sundry](#) ¹

Author affiliations

Published Feb 08, 2022 on Dryad. <https://doi.org/10.5061/dryad.7d7wm37wr>

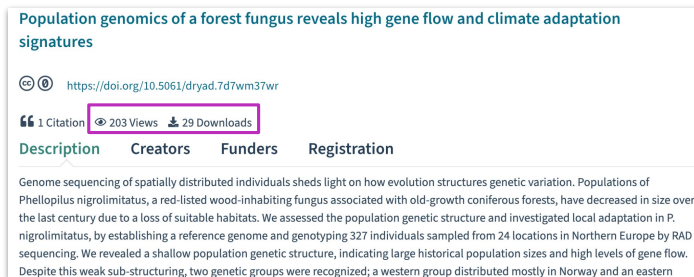
Usage Statistics: 209 views, 27 downloads, 1 citations

Share: [Twitter](#) [Facebook](#) [LinkedIn](#) [Email](#)

License: Public domain@0

Subject keywords: [Genotype-environment associations](#), [population genomics](#), [post-glacial colonisation](#), [RAD Sequencing](#), [wood](#)

Data files



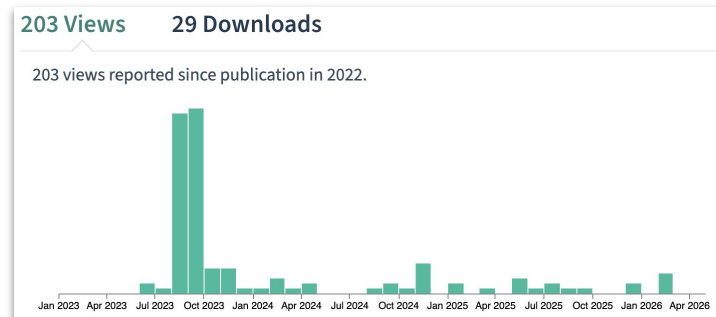
Population genomics of a forest fungus reveals high gene flow and climate adaptation signatures

<https://doi.org/10.5061/dryad.7d7wm37wr>

Usage Statistics: 1 Citation, 203 Views, 29 Downloads

Description **Creators** **Funders** **Registration**

Genome sequencing of spatially distributed individuals sheds light on how evolution structures genetic variation. Populations of *Phellipilus nigrolimitatus*, a red-listed wood-inhabiting fungus associated with old-growth coniferous forests, have decreased in size over the last century due to a loss of suitable habitats. We assessed the population genetic structure and investigated local adaptation in *P. nigrolimitatus*, by establishing a reference genome and genotyping 327 individuals sampled from 24 locations in Northern Europe by RAD sequencing. We revealed a shallow population genetic structure, indicating large historical population sizes and high levels of gene flow. Despite this weak sub-structuring, two genetic groups were recognized; a western group distributed mostly in Norway and an eastern



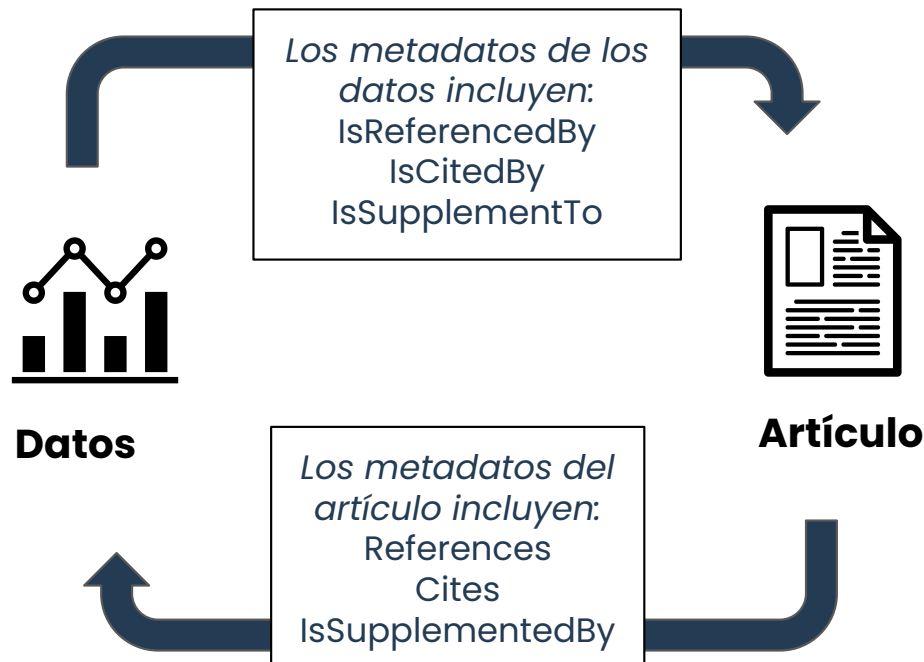
Citas de datos

La infraestructura desarrollada entre Make Data Count y DataCite permite recopilar citas de datos a través del esquema de metadatos de DataCite.

El campo **RelatedIdentifier** en el esquema de metadatos crea enlaces entre objetos.

Sub-propiedades:

- relatedIdentifierType
- relationType

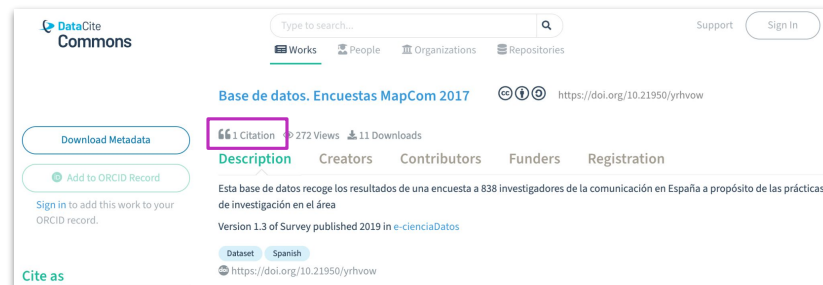


<https://schema.datacite.org/>

Citas de datos



Las citas se pueden acceder a través de los servicios de DataCite y el portal DataCite Commons.



commons.datacite.org/doi.org/10.21950/yrhvow

```
"relatedIdentifiers": [  
  {  
    "relationType": "IsSupplementTo",  
    "relatedIdentifier": "10.5209/ciyc.60910",  
    "relatedIdentifierType": "DOI"  
  }  
],
```

Lagunas en las citas de datos

Los procesos disponibles para citas de datos aún no nos dan información completa sobre la utilización de los datos de investigación.



Solo una fracción de las menciones a datos se incluyen como referencias estructuradas.



Necesidad de flujos de trabajo para capturar citas de datos con diferentes identificadores.



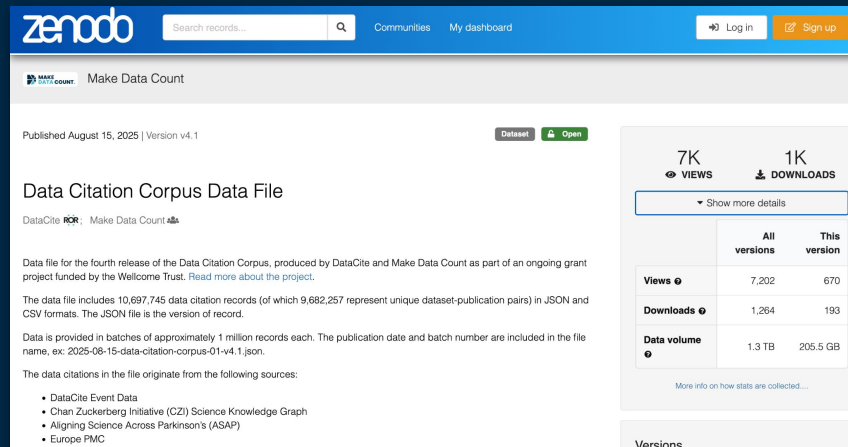
Citas identificadas por diferentes grupos se almacenan en diferentes sistemas (a veces cerrados).

Data Citation Corpus

El Corpus agrega citas identificadas mediante diferentes metodologías, incluyendo metadatos de identificadores persistentes, curación y rastreo de texto.

- Múltiples fuentes de citas: DataCite Event Data, Chan Zuckerberg Initiative, ASAP, Europe PMC
- Datos con DOIs y números de acceso

El Data Citation Corpus agrega 10 millones de citas de datos



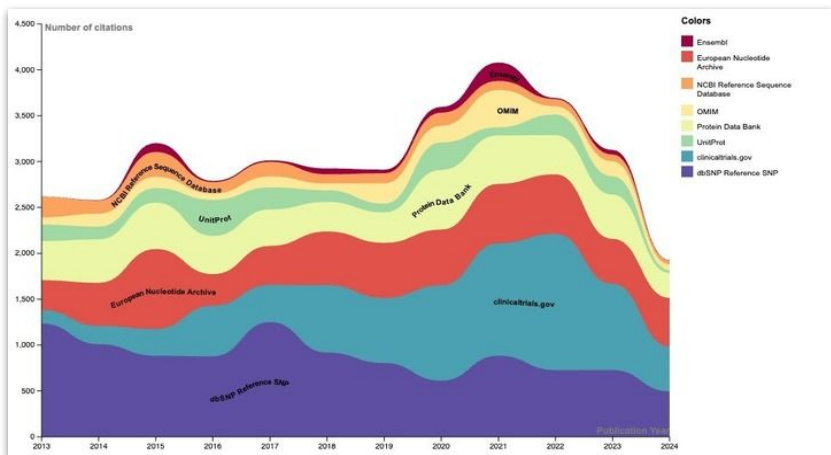
The screenshot shows the Zenodo interface for the 'Data Citation Corpus Data File'. The page includes a search bar, navigation links for 'Communities' and 'My dashboard', and a 'Log in' button. The main content area displays the file's title, publication date (August 15, 2025), and version (v4.1). It also shows the number of views (7K) and downloads (1K). A table on the right provides a breakdown of views and downloads by version. The table has three columns: 'Views', 'All versions', and 'This version'. The 'Views' row shows 7,202 views for all versions and 670 for this version. The 'Downloads' row shows 1,264 downloads for all versions and 193 for this version. The 'Data volume' row shows 1.3 TB for all versions and 205.5 GB for this version. The page also includes a description of the data file and a list of sources.

	All versions	This version
Views	7,202	670
Downloads	1,264	193
Data volume	1.3 TB	205.5 GB

Archivo del Data Citation Corpus en Zenodo:
<https://doi.org/10.5281/zenodo.11196858>



Data Citation Corpus: aplicaciones para bibliotecas



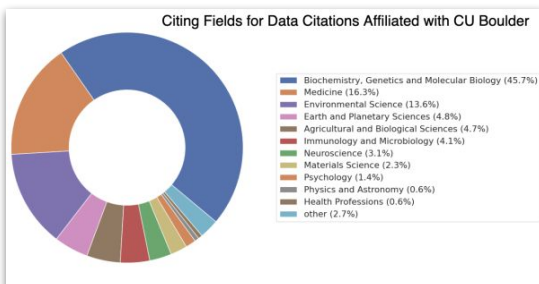
Análisis de citas de datos de las universidades de Northwestern y Colorado, Boulder, mediante citas de datos en el Data Citation Corpus y Europe PMC.

Repositorios con mayor utilización: dbSNP, Protein Data Bank, European Nucleotide Archive

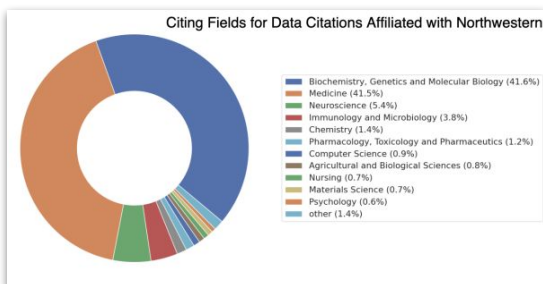
Áreas de investigación con uso intensivo de datos:

- Northwestern: cancer, inmunología, enfermedades infecciosas, biología molecular, neurociencia.
- Colorado, Boulder: ciencias ambientales, biología molecular, genética, biología vegetal.

Citing Fields for Data Citations Affiliated with CU Boulder

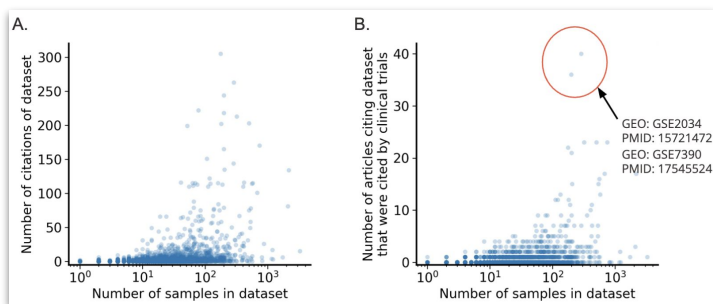


Citing Fields for Data Citations Affiliated with Northwestern



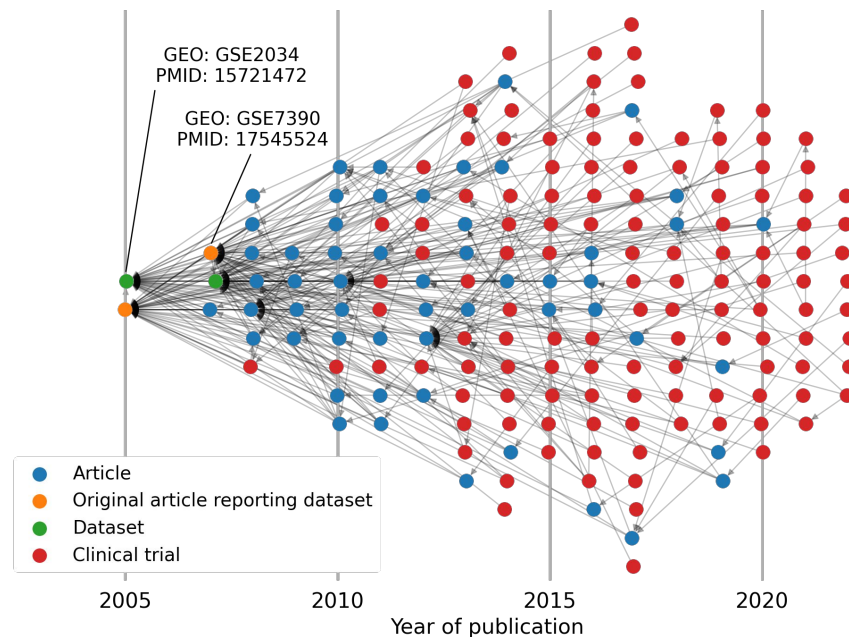
Data Citation Corpus: Explorando el impacto traslacional de los datos

El Data Citation Corpus como base para rastrear el alcance de los datos desde la ciencia básica hasta aplicaciones sanitarias.



Red de citas entre dos conjuntos de datos en GEO con gran número de citas y los artículos que describen los datos, y las publicaciones posteriores que citan esos artículos.

55 artículos citan GSE2034 y GSE7390, estos fueron posteriormente citados por 124 estudios clínicos.



Richardson, R., Puebla, I., Portenoy, J., Gutzman, K., & Holmes, K. Opening up translational data impact through the Data Citation Corpus.
<https://doi.org/10.5281/zenodo.15299655>

Los datos deben formar
parte de la **evaluación
académica**

La reforma de la evaluación debe incluir los datos abiertos

Diferentes iniciativas promueven la reforma de los sistemas de evaluación académica, con el objetivo de alinear la evaluación a principios de rigor y responsabilidad.

Uno de los ejes de esta reforma es la inclusión de un abanico amplio de resultados y actividades, incluyendo los datos.

- DORA y otras organizaciones tienen actividades en marcha para avanzar el reconocimiento a productos de investigación más allá de los artículos en revistas.
- Grupo de CoARA 'Métricas e Indicadores Responsables' está considerando los datos dentro de su trabajo.



<https://asapbio.org/international-community-convenes-in-pisa-to-advance-coordinated-reform-in-publishing-and-research-assessment/>

Grupo de Trabajo «Implementación de la Evaluación de Datos en el Ámbito Académico»



Colaboración entre Make Data Count y HELIOS Open que ha desarrollado recursos para la implementación de la evaluación de datos en procesos institucionales.

Modelo de madurez para la evaluación de datos en procesos institucionales

Herramienta de autoevaluación para que las instituciones puedan valorar sus prácticas y capacidad para evaluar los datos de investigación.

<https://makedatacount.org/explore-resources/spanish-translated-resources/>

	Inicial	Emergente	Consolidada	Optimizada
Normas y Expectativas Culturales	Los datos se toman como subproducto en lugar de considerarse como resultado significativo académico. Hay escasa o nula conciencia institucional respecto de evaluar los datos de investigación. Los datos se toman como un resultado académico sugerido, pero las expectativas siguen siendo de cumplimiento normativo e inconsecuentes. Las normas institucionales sobre el seguimiento de la evaluación y la evaluación de datos son débiles, y las prácticas están en gran medida desconectadas por personas implicadas o mandatos externos.	Comenzan a surgir conversaciones sobre la evaluación de datos como resultado académico. Algunas investigaciones y datos se evalúan los valores y objetivos institucionales. Esta sensibilización se refleja en la formalización de prácticas para seleccionar y evaluar los datos en algunas departamentos y en la integración de procedimientos fuera de los trabajos institucionales. Las normas sobre la comparación, datos y modificación de datos son cada vez más habituales, aunque las expectativas de impacto aún no son homogéneas en toda la institución.	Existe un reconocimiento creciente de que los datos son parte esencial de la cultura académica y de que la comparación y la modificación de datos afectan los valores y objetivos institucionales. Esta sensibilización se refleja en la formalización de prácticas para seleccionar y evaluar los datos en algunas departamentos y en la integración de procedimientos fuera de los trabajos institucionales. Las normas sobre la comparación, datos y modificación de datos son cada vez más habituales, aunque las expectativas de impacto aún no son homogéneas en toda la institución.	Existe un entendimiento común de que la comparación y la modificación de datos afectan los valores y objetivos institucionales. Esta sensibilización se refleja en la formalización de prácticas para seleccionar y evaluar los datos en algunas departamentos y en la integración de procedimientos fuera de los trabajos institucionales. Las normas sobre la comparación, datos y modificación de datos son cada vez más habituales, aunque las expectativas de impacto aún no son homogéneas en toda la institución.
Políticas y Procesos	Las políticas institucionales y a nivel de los departamentos de los datos se desarrollan de forma participativa y la modificación de los datos como contribuciones académicas. No existen políticas, directrices de evaluación ni flujo de trabajo para documentar o reconocer los datos como contribuciones académicas. Como resultado, las actividades relacionadas con los datos no están previstas ni se tienen en cuenta en los procesos de comparación, acciones o evaluación de rendimiento académico, y hay pocos procesos internos que apoyen su realización.	Las políticas institucionales reconocen los datos como contribuciones académicas, pero principalmente en relación con el cumplimiento normativo, la propiedad o las responsabilidades de profesores y no como resultados académicos. Existen procesos para evaluar las actividades relacionadas con los datos en los materiales de evaluación, pero son esporádicos e inconsistentes. Se apoyan con otros categorías como la divulgación o resultados de investigación en general. El reconocimiento sigue siendo inconsistente y depende en gran medida del interés individual de los evaluadores o los departamentos.	Las políticas institucionales identifican explícitamente la comparación, la modificación y el impacto de los datos como contribuciones académicas. Los procesos para documentar y reconocer los datos como contribuciones académicas están cada vez más formalizados, con directrices que exigen su inclusión en currículos, expedientes y evaluaciones de rendimiento académico. Las expectativas de informar sobre estas prácticas son cada vez más claras, aunque su aplicación sigue siendo desigual a través de departamentos y disciplinas.	Las políticas institucionales exigen la documentación y evaluación de los datos en los procesos de comparación, acciones o evaluación. Los procesos para documentar y reconocer los datos como contribuciones académicas están cada vez más formalizados, con directrices que exigen su inclusión en currículos, expedientes y evaluaciones de rendimiento académico. Las expectativas de informar sobre estas prácticas son cada vez más claras, aunque su aplicación sigue siendo desigual a través de departamentos y disciplinas.
Incentivos y Reconocimiento	Los datos no se reconocen como resultados académicos de impacto a nivel general, quedan excluidos de las evaluaciones de rendimiento académico o académicas. No existen reconocimientos institucionales o académicos relacionados con el desarrollo profesional –que apoyen o fomenten la comparación de datos o resultados de investigación. Cuando se produce algún reconocimiento, se lo da sólo en el caso de cuando alguien lo ha sido vinculado a prácticas institucionales.	Los datos reciben un reconocimiento ocasional, inconsistentemente resultado por iniciativas individuales o por requisitos específicos de agencias de financiación. Algunas departamentos ofrecen reconocimientos institucionales o académicos en evaluaciones anuales, pero estos esfuerzos son inconsistentes y fragmentarios. El reconocimiento y las recompensas por la comparación de datos son inconsistentes o ofrecen recompensas institucionales que no están vinculadas al reconocimiento de esas actividades.	La institución y muchos departamentos reconocen los datos como contribuciones académicas, y el impacto de los datos como contribuciones académicas, recomendando su inclusión en currículos, expedientes para enseñar e investigar. Existen programas de desarrollo profesional que reconocen los datos como contribuciones académicas. El reconocimiento y las recompensas se aplican de forma desigual y al reconocimiento de los datos en los informes y las comunicaciones institucionales.	Los datos y su impacto son plenamente reconocidos como contribuciones académicas y están integrados de forma consistente en los marcos de evaluación de investigadores, tanto a nivel institucional como departamental. Los programas de desarrollo profesional reconocen los datos como contribuciones académicas. El reconocimiento y las recompensas se aplican de forma desigual y al reconocimiento de los datos en los informes y las comunicaciones institucionales.
Infraestructura	Los sistemas y plataformas institucionales operan de forma aislada, sin mecanismos para conectar o coordinar las actividades de comparación de datos. Las identificaciones parciales (PIPs) no se utilizan de manera sistemática para realizar el seguimiento de investigadores, datos o resultados académicos, y no existe estructura de gobernanza que oriente la gestión o la evaluación de los datos. No hay herramientas institucionales o de datos para medir las actividades de comparación de datos. Si se aplica, el apoyo técnico dedicado a esas actividades.	Existe una integración limitada entre algunas plataformas, lo que permite un seguimiento parcial de la comparación de los datos, aunque la mayor parte de la información permanece desconectada. Algunas herramientas, como los PIPs, facilitan actividades, pero su uso no es consistente y la marca de gobernanza no existe. Existen herramientas limitadas o prácticas de cumplimiento normativo a nivel departamental. Existen herramientas o de datos para medir las actividades de comparación de datos. Si se aplica, el apoyo técnico dedicado a esas actividades.	Los sistemas y flujos de datos institucionales permiten el seguimiento regular de la comparación de los datos y la medición individual de impacto, aunque la información está fragmentada. Algunas herramientas, como los PIPs, facilitan actividades, pero su uso no es consistente y la marca de gobernanza no existe. Existen herramientas limitadas o prácticas de cumplimiento normativo a nivel departamental. Existen herramientas o de datos para medir las actividades de comparación de datos. Si se aplica, el apoyo técnico dedicado a esas actividades.	Los sistemas institucionales son completamente interoperables y están integrados a entornos institucionales, lo que permite un seguimiento regular de la comparación y la modificación de datos. Algunas herramientas, como los PIPs, facilitan actividades, pero su uso no es consistente y la marca de gobernanza no existe. Existen herramientas limitadas o prácticas de cumplimiento normativo a nivel departamental. Existen herramientas o de datos para medir las actividades de comparación de datos. Si se aplica, el apoyo técnico dedicado a esas actividades.
Desarrollo de Capacidades	No existen esfuerzos coordinados de formación o desarrollo profesional en la evaluación de datos o en la evaluación de datos. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores.	Algunas unidades ofrecen apoyo o formación limitada que se centra en la evaluación de datos o en la evaluación de datos. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores. La formación de investigadores se centra en la formación de investigadores, pero no en la formación de investigadores.	Las oportunidades de formación y divulgación de los datos están disponibles para todos los investigadores, aunque la información está fragmentada. Algunas herramientas, como los PIPs, facilitan actividades, pero su uso no es consistente y la marca de gobernanza no existe. Existen herramientas limitadas o prácticas de cumplimiento normativo a nivel departamental. Existen herramientas o de datos para medir las actividades de comparación de datos. Si se aplica, el apoyo técnico dedicado a esas actividades.	La formación sobre la evaluación de datos está integrada en la formación de investigadores, tanto a nivel institucional como departamental. Los programas de desarrollo profesional reconocen los datos como contribuciones académicas. El reconocimiento y las recompensas se aplican de forma desigual y al reconocimiento de los datos en los informes y las comunicaciones institucionales.
Alineamiento y Compromiso Externo	Existe escasa o nula conciencia institucional sobre los esfuerzos de reforma de la evaluación de la investigación, como DORA, CAHPEL, HELIOS Open, o otros esfuerzos institucionales o académicos. Las políticas y las prácticas de evaluación institucional no son proactivas para evaluar la diversidad de resultados de investigación, y no hay reconocimiento de cómo los datos pueden contribuir a los objetivos o prioridades institucionales. La institución no participa o contribuye a iniciativas externas de reforma.	La conciencia sobre iniciativas externas de reforma de la evaluación de la investigación, como DORA, CAHPEL, HELIOS Open, o otros esfuerzos institucionales o académicos. Las políticas y las prácticas de evaluación institucional no son proactivas para evaluar la diversidad de resultados de investigación, y no hay reconocimiento de cómo los datos pueden contribuir a los objetivos o prioridades institucionales. La institución no participa o contribuye a iniciativas externas de reforma.	La institución incorpora cada vez más en sus políticas y prácticas los principios de evaluación de datos, como la diversidad, la equidad y la inclusión. Las directrices institucionales y los esfuerzos de evaluación reconocen la importancia de valorar la diversidad de contribuciones académicas, incluidas las de datos, con referencia a requisitos formales o en prácticas de evaluación. La participación en iniciativas externas de reforma de la evaluación de la investigación se ha convertido en una prioridad para la institución y se ha convertido en una prioridad para la institución y se ha convertido en una prioridad para la institución.	La institución demuestra un compromiso asertivo con la reforma de la evaluación de la investigación, como DORA, CAHPEL, HELIOS Open, o otros esfuerzos institucionales o académicos. Las políticas y las prácticas de evaluación institucional no son proactivas para evaluar la diversidad de resultados de investigación, y no hay reconocimiento de cómo los datos pueden contribuir a los objetivos o prioridades institucionales. La institución no participa o contribuye a iniciativas externas de reforma.

Grupo de Trabajo «Implementación de la Evaluación de Datos en el Ámbito Académico»



Guía de implementación «Evaluación de datos en el ámbito académico»

1. Texto de muestra para políticas de titularidad y ascensos
2. Plantilla de currículum académico
3. Orientaciones para comités de evaluación y revisora/es externa/os

<https://makedatacount.org/explore-resources/spanish-translated-resources/>

Datos en formato de curriculum anotado

DataCite, & Make Data Count. (2025). Data Citation Corpus Data File (v4.1) [Conjunto de datos]. DataCite.

<https://doi.org/10.5281/zenodo.16901115>

- *Rol: coordinador del proyecto*
- *Tipo de datos: colección de registros de metadatos*
- *Disponibilidad: licencia CC0*
- *Uso y reutilización: las diferentes versiones han tenido 6.900 visualizaciones y han sido descargadas más de 1.200 veces.*
- *Contribución: el conjunto de datos proporciona la colección más amplia disponible de citas de datos (10 millones), reuniendo en una única colección abierta las citas de diferentes fuentes y para diferentes tipos de identificadores de datos.*

Grupo de Trabajo «Implementación de la Evaluación de Datos en el Ámbito Académico»

Estudios de casos



Procesos institucionales que reconocen los datos: abriendo nuevos caminos hacia la titularidad y ascensos

Michael Dougherty
Departamento de
Psicología, Universidad
de Maryland y
HELIOS Open



Case Study

Emprendiendo el cambio cultural para ampliar los criterios de evaluación dentro de las instituciones

Christian Blanco
Subdirector de Asuntos de
Investigación
Internacional, Pontificia
Universidad Católica de
Chile



Estudio de caso

Avanzando la compartición y la evaluación de datos en una fundación de investigación

Vanessa Jorge
Fundación Oswaldo
Cruz -
FIOCRUZ, Brazil



Estudio de caso

Avanzando la utilización responsable de datos y el reconocimiento de la gestión de datos

Denise M. Scholtens
Profesora de Bioestadística e
Informática en la Facultad de
Medicina Feinberg y Directora del
Centro de Análisis y Coordinación
de Datos (NUDACC) de la
Universidad Northwestern.



Estudio de caso

Desarrollando flujos de trabajo para la publicación y reutilización de datos

Patrick Charbonneau
Catedrático de Física,
Duke University



Estudio de caso

Maneras prácticas de incorporar los datos en los procesos de titularidad y ascensos

Jeffrey Blume
Catedrático de la
Facultad de Ciencia de
Datos, Universidad de
Virginia



Estudio de caso

**¿Interesado/a en compartir la experiencia en tu universidad?
Contáctanos, invitamos casos de estudio adicionales**

Habilitar evaluaciones
diversas y responsables de
los datos

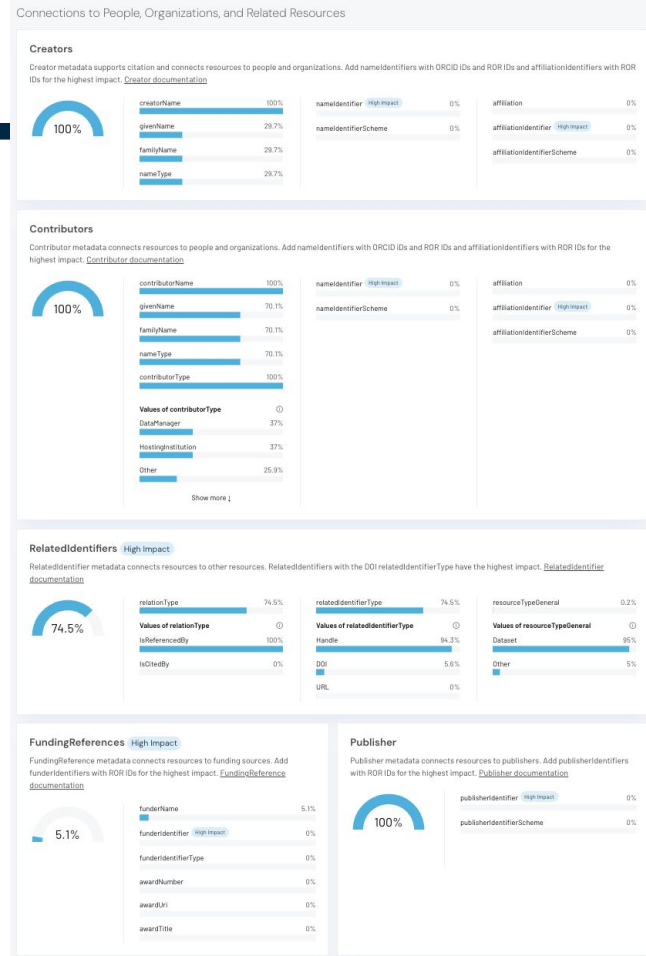
Metadatos

Metadatos completos y abiertos

- Maximizar la visibilidad.
- Hacer posible evaluaciones a nivel de agencia de financiación, institución, proyecto y otras.



<https://metadata.datacite.org/csic.digital?resourceType=dataset>

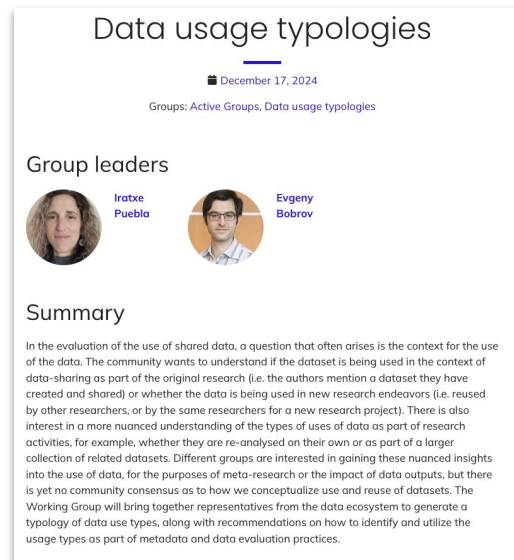


El contexto: Clave para habilitar evaluaciones diversas

El grupo de trabajo de FORCE11 'Tipología de uso de datos' reúne a diferentes actores en el ecosistema de datos con el objetivo de desarrollar una tipología común de la utilización de datos.

- Lista codificada de tipos de utilización de datos, con sus características asociadas.
- Recomendaciones para capturar esta información en metadatos.

<https://force11.org/group/data-usage-typologies/>



The screenshot shows the 'Data usage typologies' webpage. At the top, the title 'Data usage typologies' is displayed in a large, dark font. Below it, a date 'December 17, 2024' is shown next to a calendar icon. Underneath the date, the text 'Groups: Active Groups, Data usage typologies' is visible. The 'Group leaders' section features two circular profile pictures: one of Iratxe Puebla on the left and one of Evgeny Bobrov on the right. Below the 'Summary' heading, there is a paragraph of text explaining the purpose of the typology, which is to provide a common framework for understanding data use in research. The text mentions that the typology is intended to help researchers understand the context of data use, such as whether it is for original research, reanalysis, or reuse, and to provide recommendations for how to capture this information in metadata.

Tipología de uso de datos

Tres dimensiones: Acceso, Usuario, Aplicación

Access/How is use enabled? Definition of "Access": The permission and ability to retrieve, modify, copy, and/or move data.		
CATEGORY	DEFINITION	VALUES
Access: Actor	Type of entity which accesses data.	- Human
	Which type of entity accesses the data?	- Machine
Access: Data discovery	Infrastructure, technology or protocol used to establish the existence and location of data	- Search Engine
	How is the data discovered? More than one option can be chosen.	- Text-based publication - Catalog - Direct link - Identifier - Related links or identifiers - OAI-PMH - FAIR signposting - Personal communication - Other
Access: Point of Access	Provider, medium or source from which data were drawn for the instance of use	- Data Repository
	Where does the data live?	- Journal - Cloud storage - Webpage - Local medium - Other
Access: Data interaction	Technical pathway or platform through which data is transferred or interacted with for the purpose of use.	- Web download
	Which way is the data transferred or interacted with? More than one option can be chosen.	- Direct download - Email - File-hosting service - Distributed Data Storage - Secure Data Enclave - API - Manual transcription - Other
Access: Restriction on use	Applicable type(s) of non-technical (e.g. ethical, legal) restrictions on use	- License type
	Which non-technical restrictions apply to use? More than one option can be chosen.	- Local context - Data use agreement - Ethical restrictions - Committee clearance - None - Other

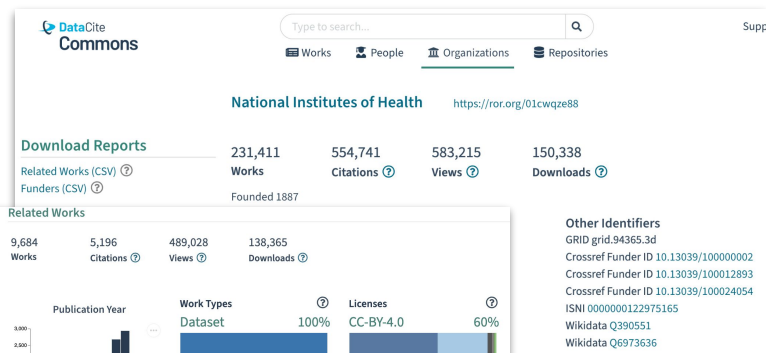
User/Who is using the data? Definition of "User": An entity (human or machine) which interacts with the data. (see relational definition of data)		
CATEGORY	DEFINITION	VALUES
User: Role	Function assumed by the user with respect to the data.	- Creator
	In which role(s) does the user engage with the data?	- Consumer - Curator - Reviewer
User: Permission	The type of authorization the user has with regard to interacting with the data.	- Administrator
	Which level of authorization does the user have?	- Editor - Viewer with download - Viewer without download - None
User: Geography	The user's geographic location	- Country
	Where is the user located?	
User: Discipline	The user's field of research	- OECD Fields of Science and Technology vocabulary
	What is/are the user's field(s) of research, if applicable? More than one option can be chosen.	- N/A

Application/How are the datasets used? Definition of "Application": Way in which the data is put to a desired use.		
CATEGORY	DEFINITION	VALUES
Application: Purpose	The scholarly, public, or private sector(s) or sphere(s) that characterize the activities including an instance of use.	- Education - Reporting - Software or tool development - Commercial product - Policy - Research - Operational analytics - Citizen science - Personal/recreational - Journalism
	Why is the user interacting with the data? More than one option can be chosen.	- Review - Curation - Other
Application: Activity type	The specific actions, be them analytical, exploratory, or otherwise, leveraging data that constitute an instance of use.	- Model Development/ Training - Meta-analysis - Statistical methods - Comparison or control - Benchmarking - Exploration - Validation - Reproduction/ re-analysis - Curation/cleaning - Annotation/labeling - Review - Infrastructure - Data assimilation - Other
	How is the user leveraging or interacting with the data? More than one option can be chosen.	
Application: Time of use	Time point at which an activity related to the data occurred.	- Use start time
	When does the data use occur?	- Use end time
Application: Output	A product resulting from an instance of use.	- Controlled vocabulary
	What is produced from a given instance of use, if applicable? More than one option can be chosen.	

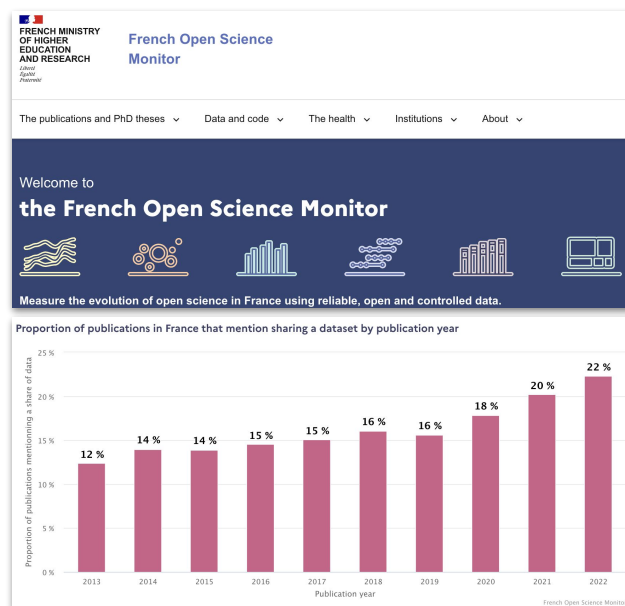
Interés creciente en comprender la utilización de datos y su contexto

Interés por agencias de financiación y a nivel nacional.

Dashboard de NIH en DataCite Commons para datos compartidos en repositorios generalistas.



Monitor Francés de Ciencia Abierta



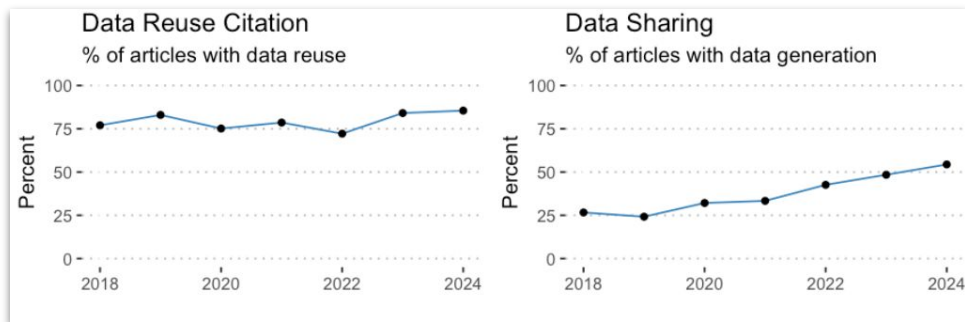
<https://frenchopensciencemonitor.esr.gouv.fr/>

Interés creciente en comprender la utilización de datos y su contexto

Se están desarrollando LLMs para identificar la utilización de datos y su contexto: utilización por l@s creadores u otr@s.

Análisis de DataSeer de 6000 publicaciones financiadas por la Fundación Michael J. Fox. Rastreo de texto para identificar la utilización de datos y si estos han sido creados para el estudio o reutilizados.

43% de los artículos muestran reutilización de datos.



Competición Kaggle de Make Data Count para desarrollar este tipo de modelos para identificar y contextualizar citas de datos.

<https://dataseer.ai/2026/03/12/uncovering-hidden-data-reuse-new-ai-technology-from-dataseer-improves-the-ability-to-measure-the-downstream-impact-of-research-data/>

La influencia de la inteligencia artificial

Oportunidades

- Enriquecimiento de metadatos
- Identificación de la utilización de datos a mayor escala
- Contextualización de la utilización y la relación entre objetos



Desafíos

- Limitaciones y sesgos de las metodologías de LLMs
- Riesgo de malas prácticas y abusos, tensión sobre la infraestructura abierta
- Indicadores y métricas para capturar la utilización mediante herramientas de IA

Infraestructura: Mejores
metadatos, interoperabilidad

Mayor visibilidad sobre la
utilización de datos y su contexto

Mayor reconocimiento a los datos
en procesos institucionales



Gracias

info@makedatacount.org

makedatacount.org



MAKE
DATA COUNT.

Del acceso al impacto: Claves para evaluar los datos de investigación

Iratxe Puebla, Make Data Count
XXI Workshop de Proyectos Digitales
28 de mayo del 2026



MAKE
DATA COUNT.