

Taller GBIF.ES: Estandarización, documentación y publicación de datos de seguimientos de biodiversidad

28 – 30 de octubre de 2025, Universidad de Sevilla

Introducción a la publicación de datos de biodiversidad

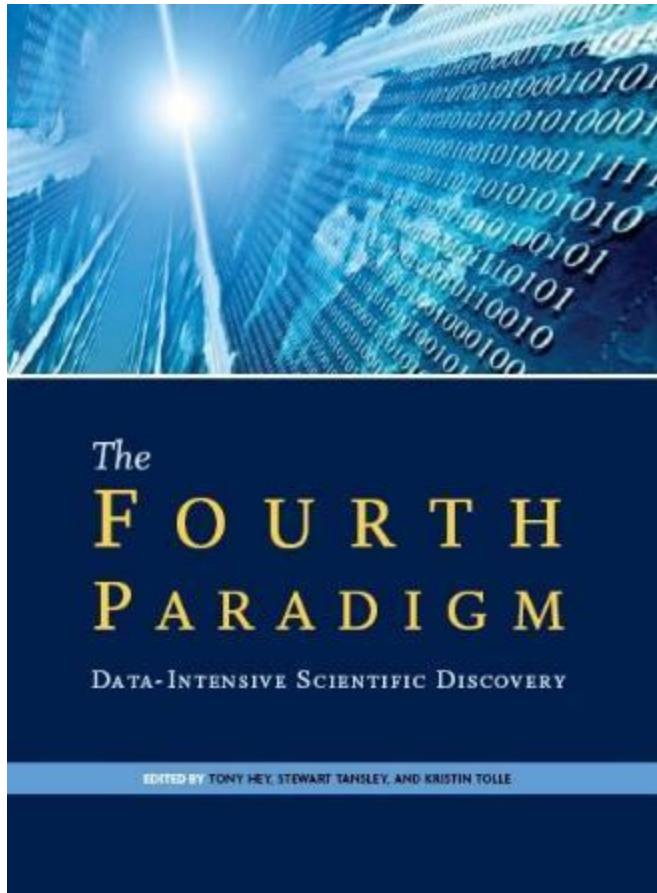
Francisco Pando
GBIF España



Qué vamos a ver

- La importancia de los datos en ciencia
- La importancia de publicar los datos
- La importancia de publicar los datos de biodiversidad
- El paradigma en el que estamos: ciencia abierta
- Ciencia abierta y GBIF

Datos y Ciencia: el 4º paradigma



Science Paradigms

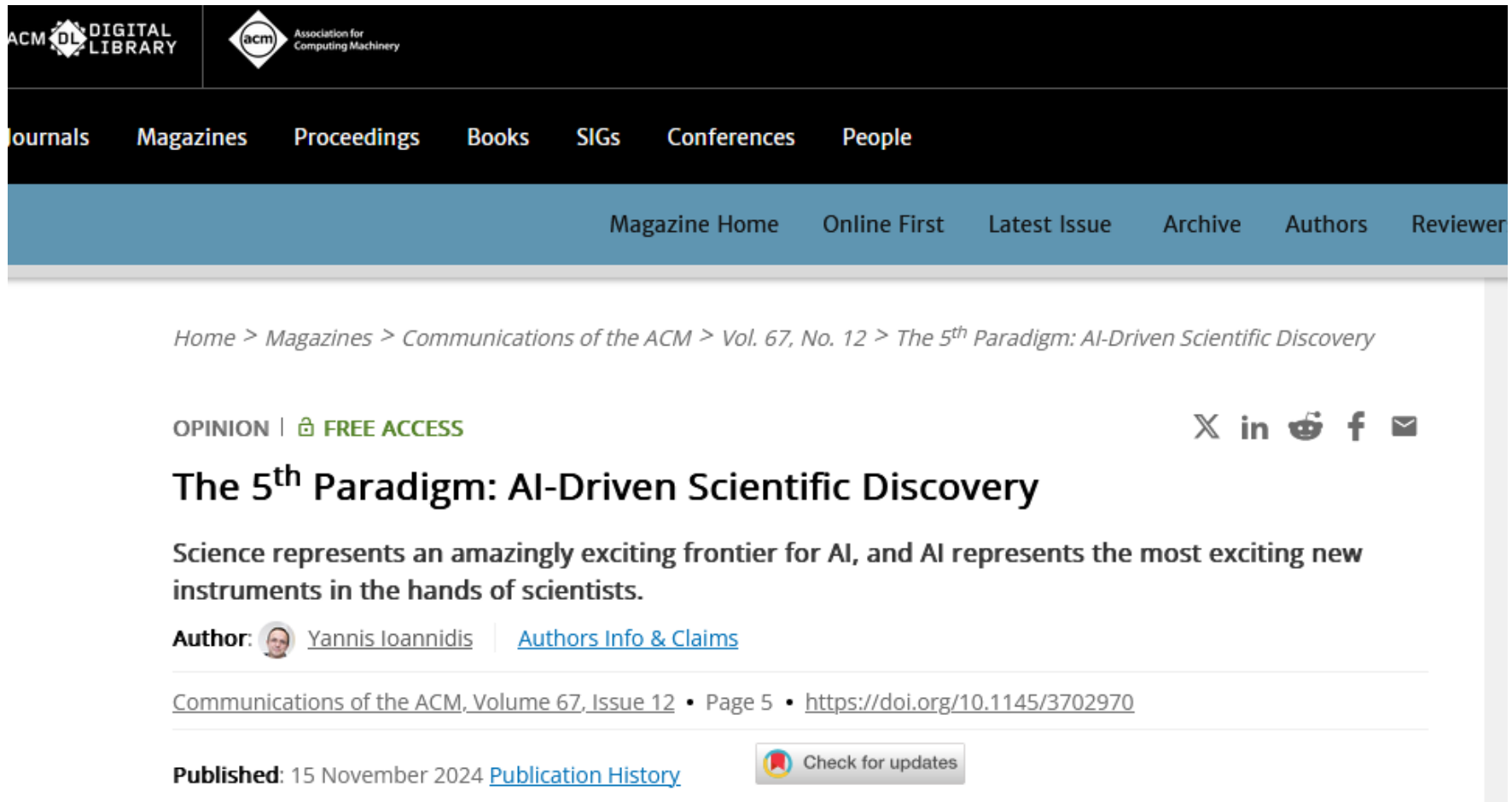
- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



Datos y Ciencia: el 5º paradigma

- *The 5th Paradigm: AI-Driven Scientific Discovery*









ACM DIGITAL LIBRARY Association for Computing Machinery

Journals Magazines Proceedings Books SIGs Conferences People


Magazine Home Online First Latest Issue Archive Authors Reviewer

Home > Magazines > Communications of the ACM > Vol. 67, No. 12 > The 5th Paradigm: AI-Driven Scientific Discovery


OPINION |  FREE ACCESS     

The 5th Paradigm: AI-Driven Scientific Discovery

Science represents an amazingly exciting frontier for AI, and AI represents the most exciting new instruments in the hands of scientists.

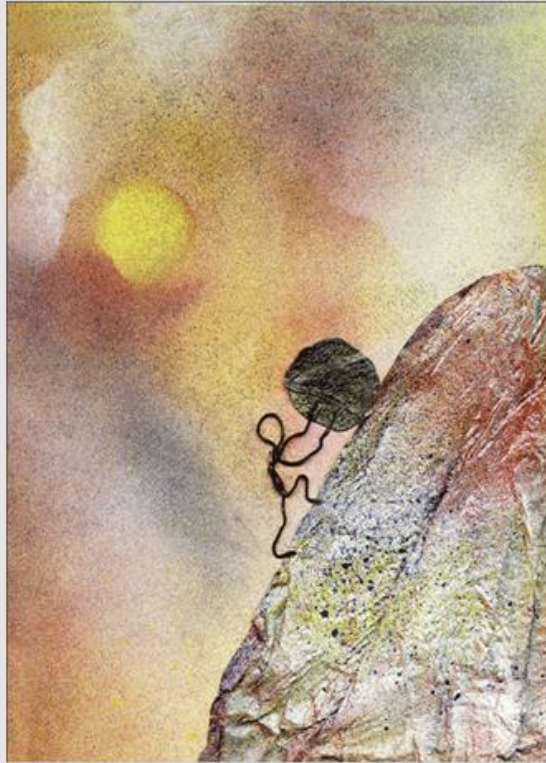
Author:  Yannis Ioannidis [Authors Info & Claims](#)

Communications of the ACM, Volume 67, Issue 12 • Page 5 • <https://doi.org/10.1145/3702970>

Published: 15 November 2024 [Publication History](#) 

<https://dl.acm.org/doi/10.1145/3702970>

Porqué Ciencia abierta: Sísifo, ciencia e informática



© The Art of
Steven Nelson

HEYWOOD, V.H. (1974). Systematics-the stone of Sisyphus. Biol. J. Linn. Soc. 6(2): 169-178.

El acceso abierto a las publicaciones (artículos, libros), mejora la eficiencia

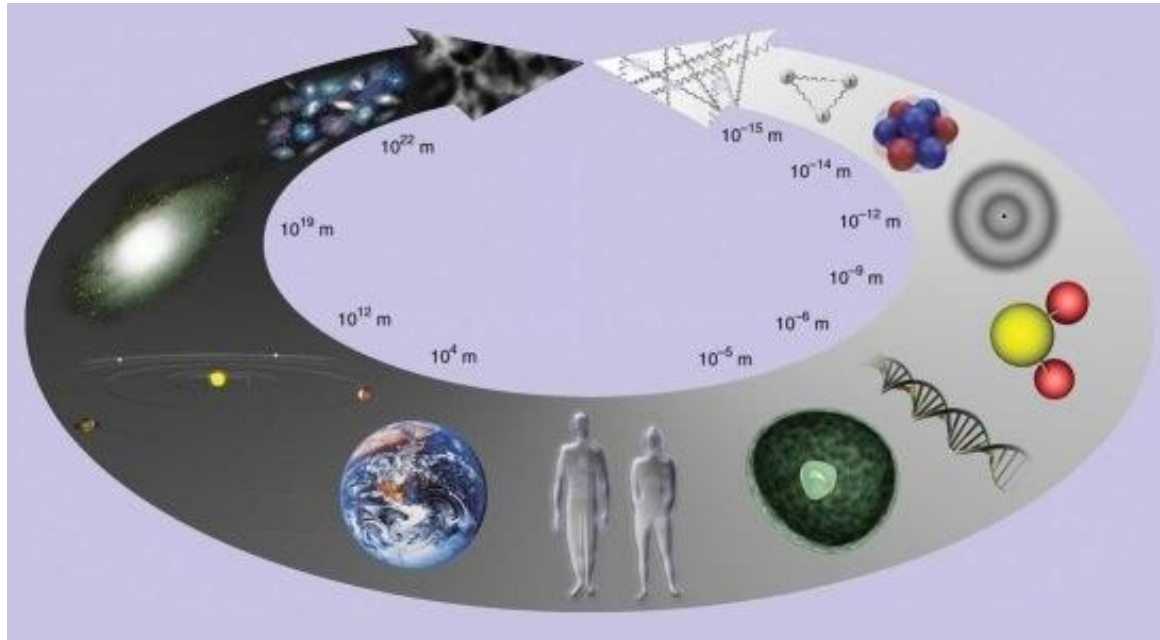
El acceso abierto a los datos lo cambia todo:

- Reproducibilidad y transparencia
- Verificación y confianza pública
- Velocidad e innovación
- Colaboración y multidisciplinariedad
- Mayor impacto y citación

Metodologías abiertas:

- Código abierto
- Entornos reproducibles

“Los **datos de Biodiversidad** rara vez son puros y nunca son sencillos”



- Los desafíos globales son –directa o indirectamente– desafíos de la biodiversidad
- Todos somos partes interesadas en la biodiversidad
- ⇒ Los datos de biodiversidad se recopilan de múltiples maneras, bajo innumerables restricciones y requerimientos
- ⇒ Los datos de biodiversidad se utilizan de innumerables maneras



Restos ambientales: que son, porqué importan

1. Cambio climático

¿Por qué es urgente?

- Está alterando el equilibrio del sistema climático global.
- Tiene **efectos acumulativos e irreversibles** a largo plazo.

Retos asociados:

- Aumento de temperatura global (ya por encima de +1,1 °C respecto a niveles preindustriales).
- Derretimiento de glaciares y polos.
- Aumento del nivel del mar.
- Eventos extremos más frecuentes: sequías, olas de calor, huracanes, inundaciones.
- Pérdida de productividad agrícola, inseguridad alimentaria.
- Migraciones climáticas y conflictos socioambientales.

3. Contaminación (aire, agua y suelo)

4. Crisis hídrica y gestión del agua

5. Contaminación por plásticos y residuos emergentes

2. Pérdida de biodiversidad y colapso ecosistémico ¿Por qué es urgente?

- Se calcula que más de **1 millón de especies** están en **peligro de extinción** (IPBES).
- La biodiversidad es **clave para el funcionamiento de los ecosistemas** que sostienen la vida humana.

□ **Causas:**

- Destrucción y fragmentación de hábitats.
- Sobreexplotación de recursos naturales.
- Especies invasoras, Contaminación, Cambio climático.

□ **Consecuencias:**

- Colapso de servicios ecosistémicos (polinización, regulación hídrica, ciclo de nutrientes).
- Inseguridad alimentaria y sanitaria.
- Mayor riesgo de pandemias zoonóticas.

6. Urbanización descontrolada y pérdida de suelo fértil

7. Sistemas agroalimentarios insostenibles

8. Injusticia ambiental y desigualdad climática

La Ciencia abierta es la que se puede reutilizar



<https://www.flickr.com/photos/kaptainkobold/3203311346>

- Por personas y por máquinas
(entornos de datos masivos, IA, etc.)
- No basta con que esté accesible:
hace falta **permisos, contexto y calidad**
- Accesibilidad y disponibilidad
("accessibility" & "availability")
-

Elementos que hacen la ciencia abierta

- Metadatos
- Estándares
 - De datos
 - De empaquetamiento de datos
- Identificadores
- Licencias
- Citación
- El cómo se sirve el dato

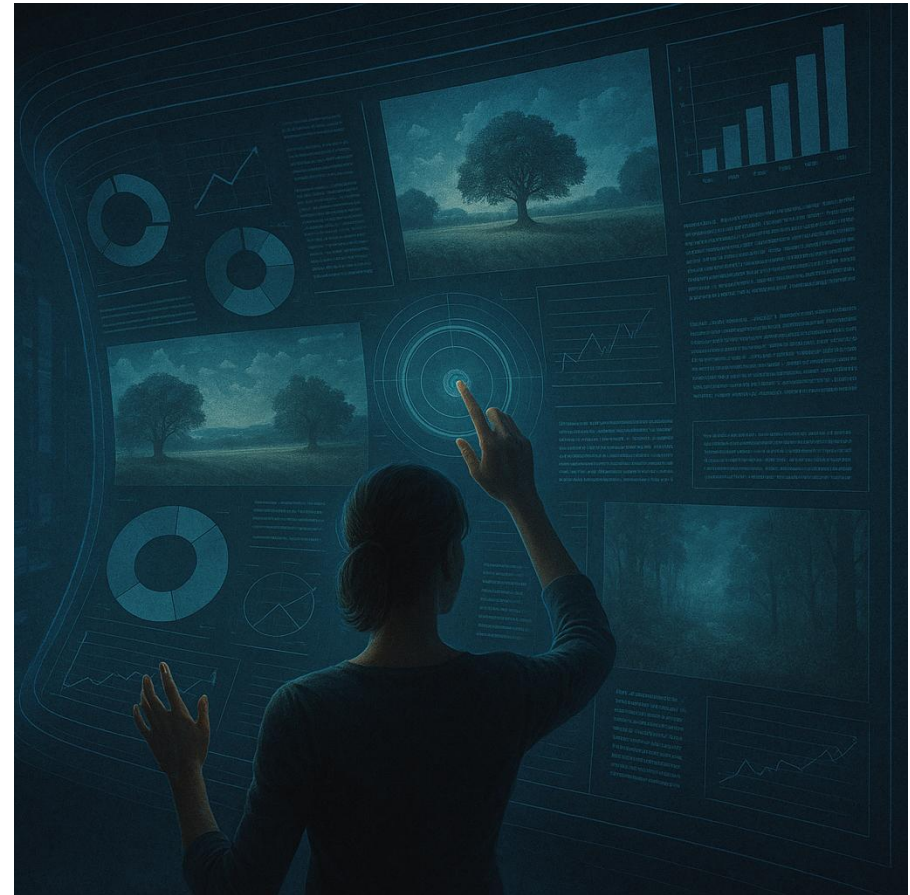


Imagen generada
con IA

Metadatos (la parte del contexto)

FORBES > INNOVATION

Data Is Worthless Without Context



Adam Blue Forbes Councils Member

Forbes Technology Council COUNCIL POST | Membership (Fee-Based)

<https://www.forbes.com/councils/forbestechcouncil/2022/01/03/data-is-worthless-without-context/>

Data, without context, is essentially useless.
Raw data lacks meaning until processed and organized into information, which then allows for insights to be drawn and decisions made

Datos sin contexto no son datos

1. ¿Quién?:

1. ¿Quién creó los datos?
2. ¿Quién es responsable de mantener los datos?

2. ¿Qué?:

1. ¿Cuál es el contenido o tema de los datos?
2. ¿En qué formato están los datos?

3. ¿Cuándo?:

1. ¿Cuándo se crearon los datos?
2. ¿Cuándo se actualizaron o modificaron por última vez?

4. ¿Dónde?:

1. ¿Dónde se recopilaron o crearon los datos?
2. ¿Dónde están almacenados los datos?

5. ¿Por qué?:

1. ¿Por qué se crearon los datos?
2. ¿Por qué son importantes o útiles?

6. ¿Cómo?:

1. ¿Cómo se recopilaron o generaron los datos?
2. ¿Cómo están estructurados y organizados los datos?

7. Calidad de los datos:

1. ¿Son precisos y fiables los datos?
2. ¿Cuáles son las limitaciones y problemas con los datos?

8. Acceso y uso:

1. ¿Quién puede acceder a los datos?
2. ¿Existen restricciones en el uso de los datos?

9. Relaciones:

1. ¿Cómo se relacionan estos datos con otros de datos?
2. ¿Hay enlaces a información adicional relevante?

Estándares

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



- Si lo cambias, ya no es un estándar
- Si solo lo usas tu, no es un estándar
- Los estándares son dinámicos y están, en constante evolución: revisiones y ediciones

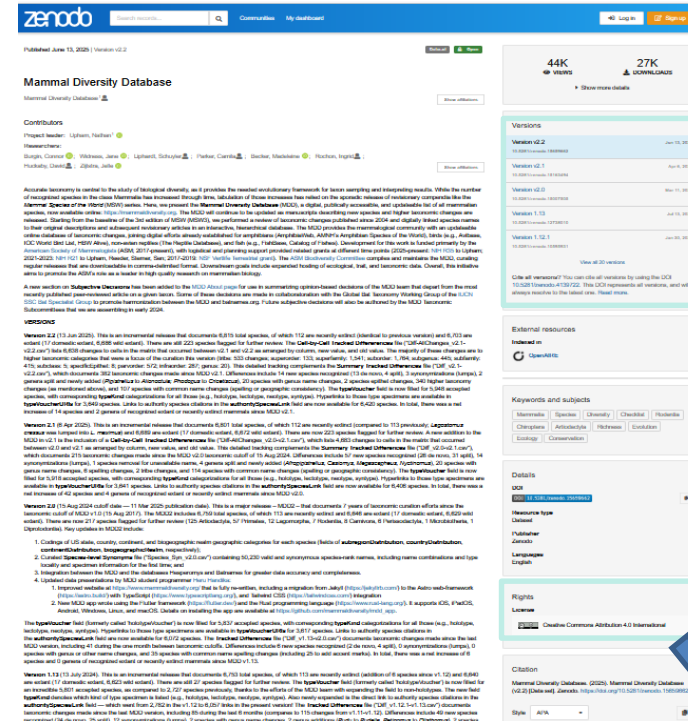
<https://xkcd.com/927/>

“pensaba que después del COVID no tendría que contar estas cosas...”

Citación

Por qué es importante incluir un apartado “Cómo citar este recurso”:

- No es obvio en muchas ocasiones
- Ayuda a garantizar el Crédito y atribución
- Facilita la trazabilidad
- Proporciona consistencia



<https://doi.org/10.5281/zenodo.156>

- Incluye un “PID”
- Ojo a las versiones (cómo citar cada versión, como citar la última)

Versions

Version v2.2	Jun 13, 2025
10.5281/zenodo.15659662	
Version v2.1	Apr 6, 2025
10.5281/zenodo.15163494	
Version v2.0	Mar 11, 2025
10.5281/zenodo.15007505	
Version 1.13	Jul 13, 2024
10.5281/zenodo.12738010	
Version 1.12.1	Jan 30, 2024
10.5281/zenodo.10595931	

[View all 20 versions](#)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.4139722](https://doi.org/10.5281/zenodo.4139722). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Citation

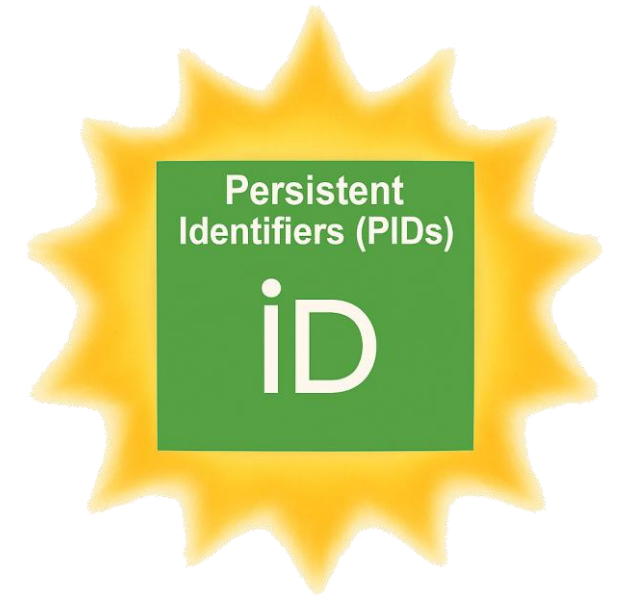
Mammal Diversity Database. (2025). Mammal Diversity Database (v2.2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15659662>

Identificadores

- Útiles para el usuario: transparencia, credibilidad, gestión, eficiencia.... Y por ello los venimos usando “desde siempre”: firmas de libros, nº D.N.I., entradas de cine, ... nombre y apellidos.
- Ahora en el mundo de la hiperconectividad, y del “machine to machine”, les pedimos mucho más, y eso conlleva alguna servidumbre.
- **Únicos, universales, persistentes y resolvibles (y opacos)**

P.ej.: Los CPR de Dinamarca

Direcciones MAC



550e8400-e29b-41d4-a716-446655440000

Adaptador de Ethernet Ethernet:

Sufijo DNS específico para la conexión. . : rjb.local
Descripción : Realtek PCIe GbE Family Controller
Dirección física. : FC-34-97-BC-A3-9A
DHCP habilitado : sí

Free and open access to biodiversity data

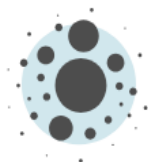
OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES



3,534,129,908

Occurrence records



118,451

Datasets



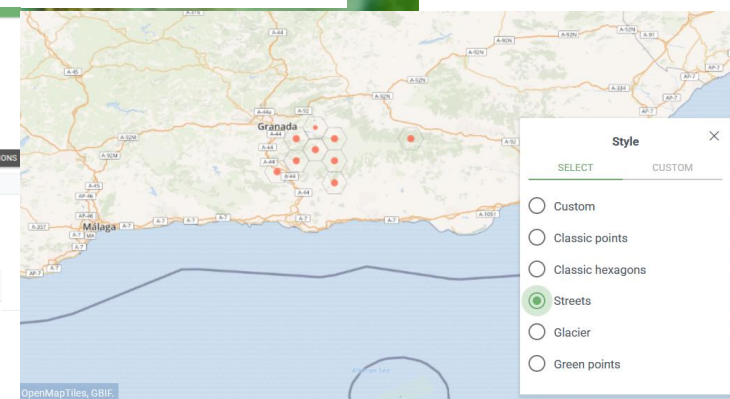
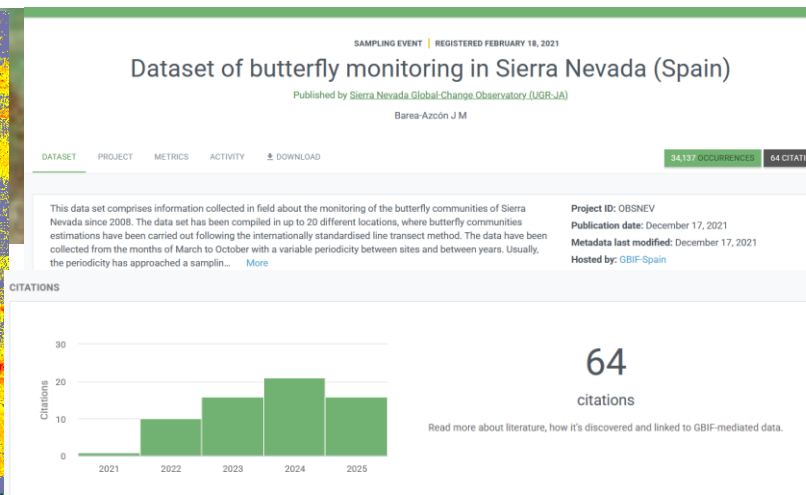
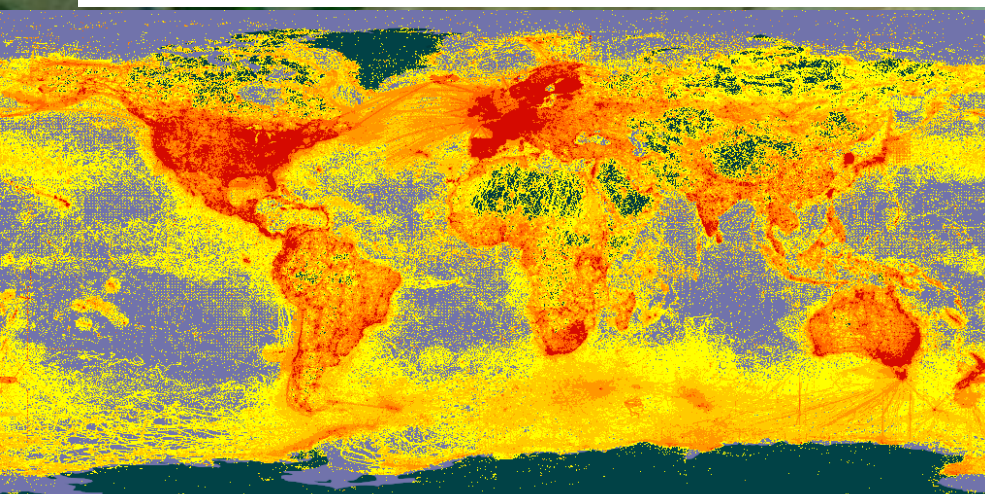
2,575

Publishing institutions



13,712

Peer-reviewed papers
using data



Qué se publica en GBIF

<https://data-blog.gbif.org/post/choose-dataset-type/>

Metadata only dataset

No data content required.
You know what is in your collection and you can describe its content and scope but you cannot make the data content available on GBIF.

Checklist

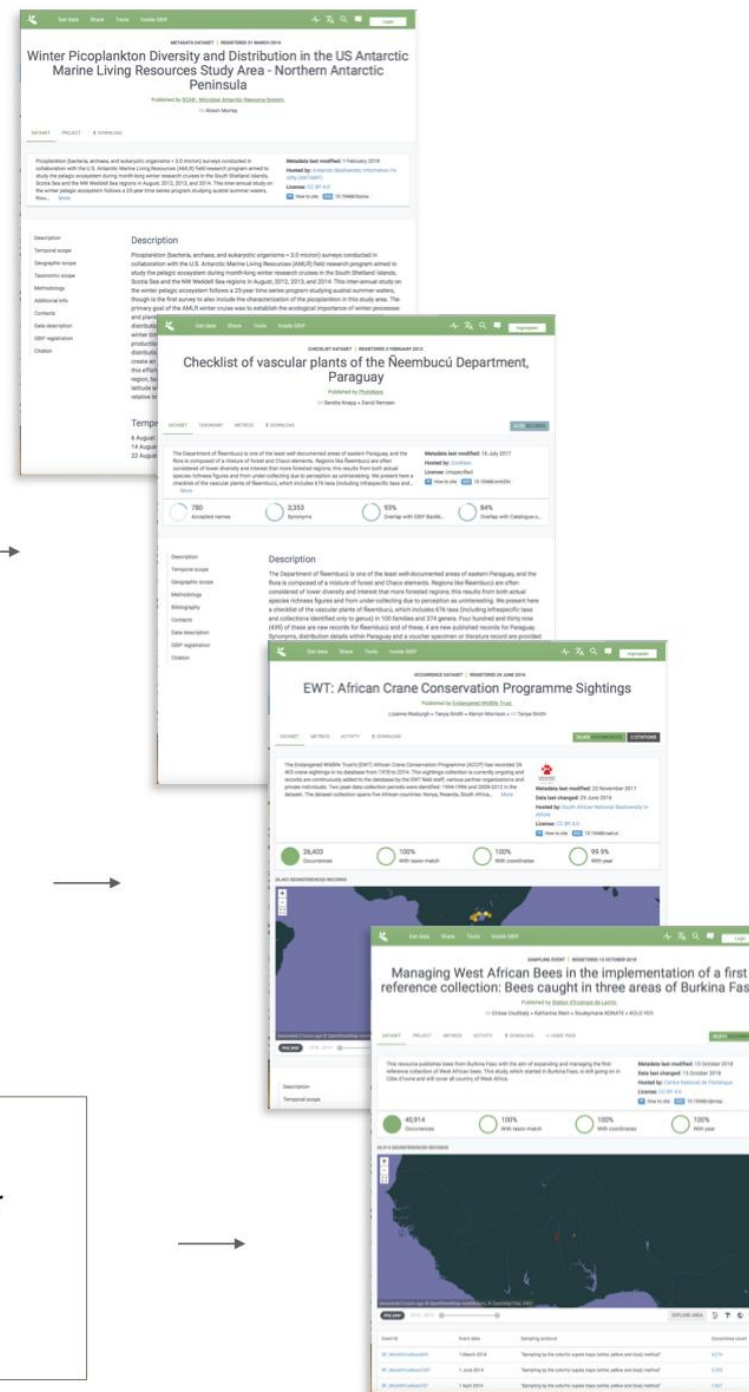
- Scientific names of organisms sharing a common theme or feature (for example: medicinal use).

Occurrence Dataset

- Scientific names of organisms observed or specimens collected,
- Observation or sampling date (year),
- Observation or sampling location (at least country).

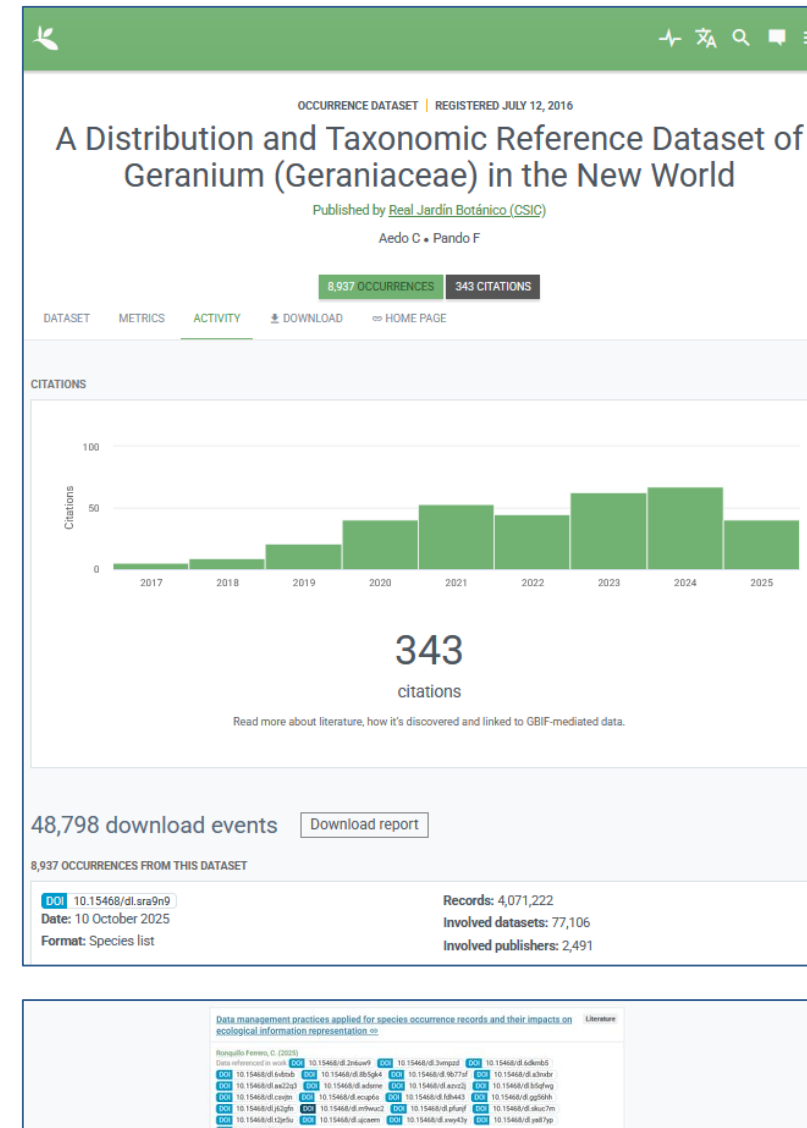
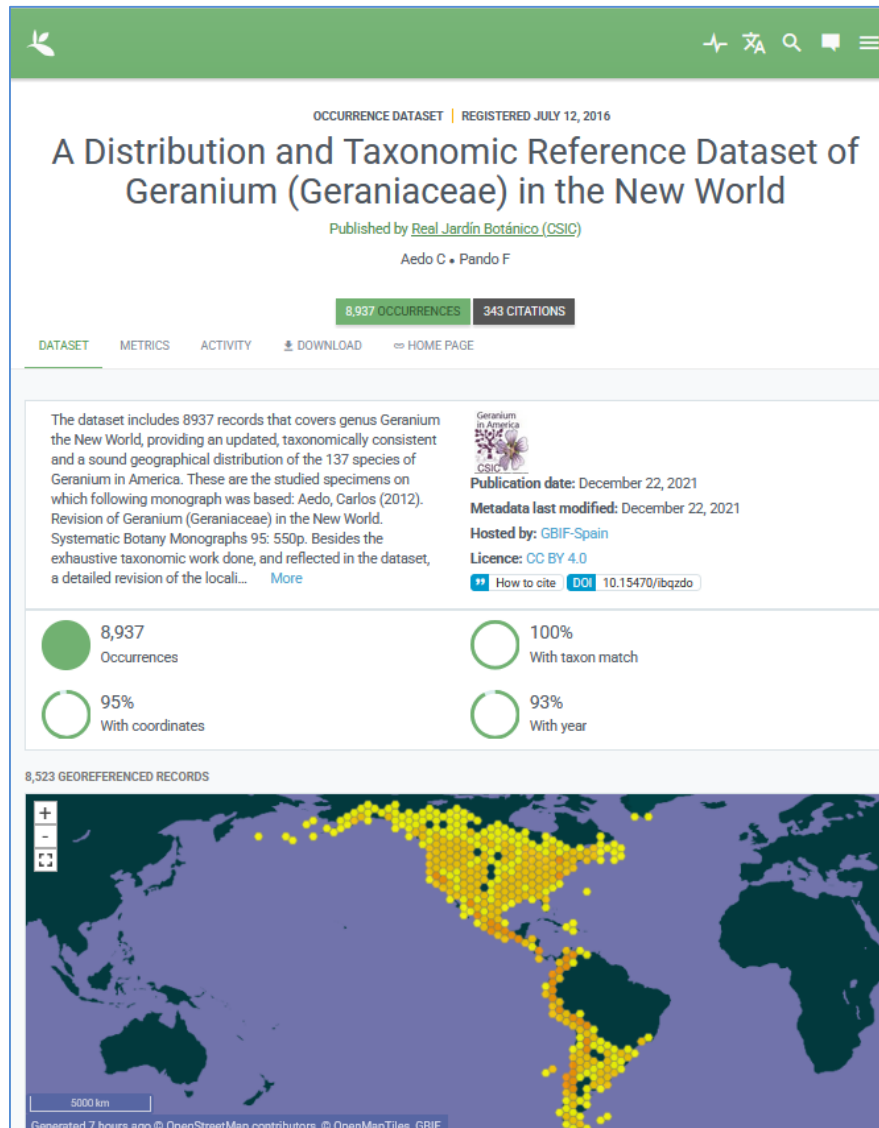
Sampling-Event Dataset

- Scientific names of organisms observed or specimens collected,
- Sampling date,
- Observation or sampling location,
- Sampling protocol.



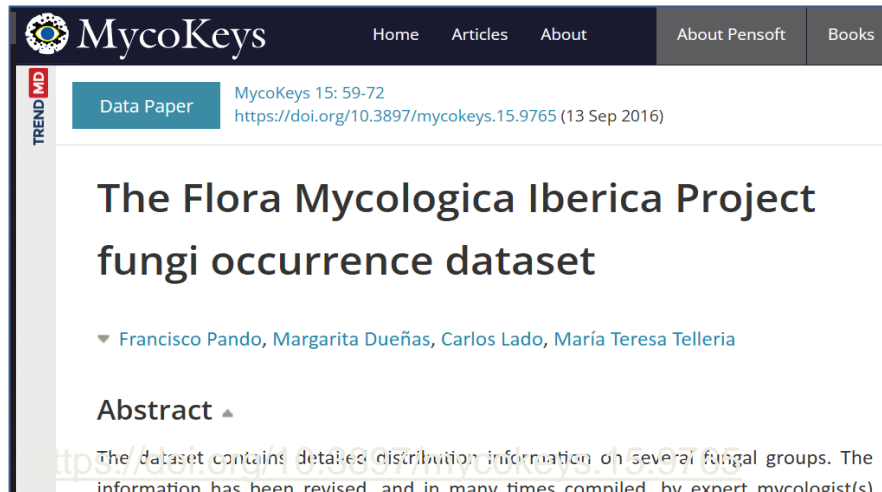
¿Que cambia los DOIs?

(Identificadores únicos, universales, persistentes y resolvibles)



<https://doi.org/10.15470/ibqzdo>

Identificadores persistentes, únicos, **resolvibles**: DOIs



DOI: Digital Object Identifier

- Ámplia implantación
- Robusto
- Resolvable

trazabilidad

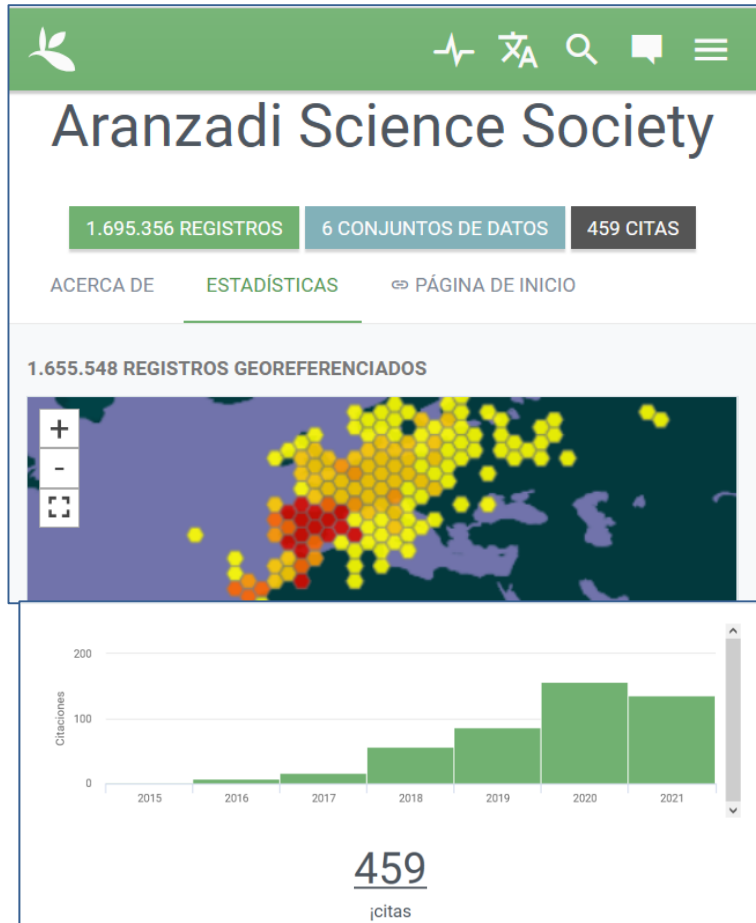
Citación ↔ crédito

Su uso se amplia más allá de las publicaciones: juegos de datos, gráficos, imágenes, descargas...

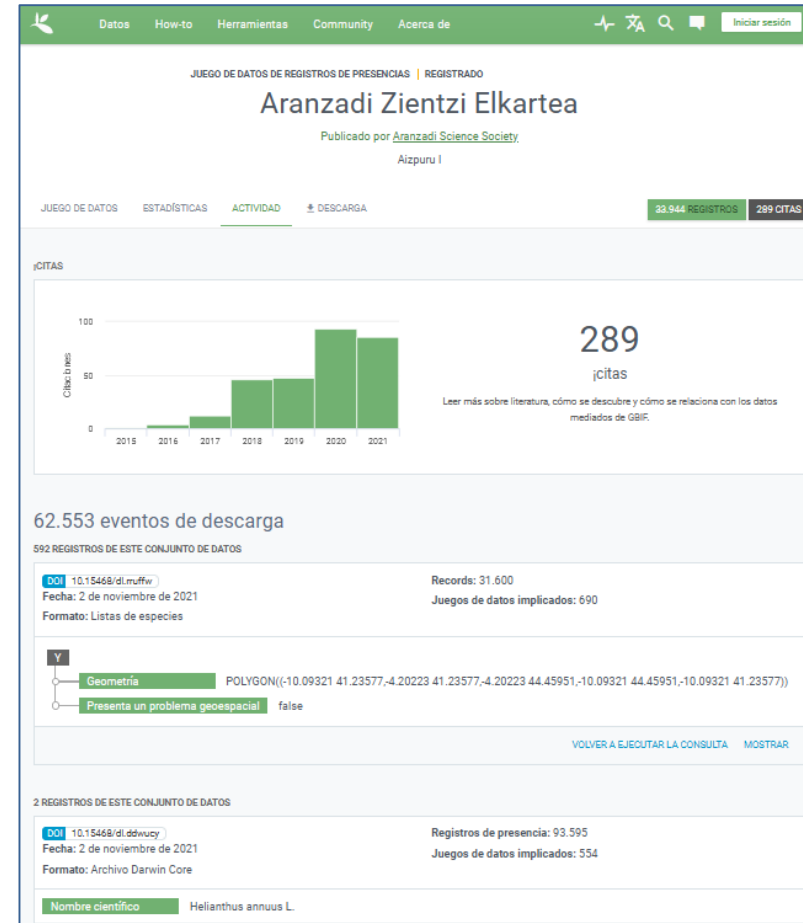
- **centralizado**
- **de pago**

¿Que cambia los DOIs?

(Identificadores únicos, universales, persistentes y resolvibles)



<https://www.gbif.org/es/publisher/823818f6-0696-4e29-bc4c-b6f3817535a7/metrics>



<https://www.gbif.org/es/dataset/8354ecee-f762-11e1-a439-00145eb45e9a/activity>

¿Que cambia los DOIs?

(Identificadores únicos, universales, persistentes y resolvibles)

BUSCAR RECURSOS | 289 RESULTADOS

TODOS LITERATURA

Leer más sobre literatura, cómo se descubre y cómo se relaciona con los datos mediados de GBIF.

Ecological niche modeling to assessment of potential distribution of *Neodiprion abietis* (Harris, 1841) (Insecta, Hymenoptera, Diprionidae) in Eurasia Literatura

Grebennikov, K. (2021) *International Journal of Agricultural Sciences and Technology*
In the article first assesses the potential distribution in Eurasia of *Neodiprion abietis* (Harris, 1841) first time assessed. The species is widely distributed in North America fir and spruce defoliator, intercepted in 2016 in the Netherlands. Analysis of the literature data on the known distribution...

Conifers • Forest pests • *Neodiprion abietis* • Niche modeling • Potential distribution

Artículo de revista Revisado por pares

Datos usados en el estudio [DOI 10.15468/dl.cdtm5m](#) [DOI 10.15468/dl.eesrd5](#) [DOI 10.15468/dl.t4c9sx](#)

Tropical-temperate dichotomy falls apart in the Asian Palmate Group of Araliaceae Literatura

Coca-de-la-Iglesia, M. Medina, N. Wen, J. Valcárcel, V. (2021) *bioRxiv*
PREMISE The use of climatic data on phylogenetic studies has greatly increased in the last decades. High-quality spatial data and accurate climatic information are essential to minimize errors in the climatic reconstructions to the past. However, despite the huge amount of already available biodiversity...

Documento de trabajo Acceso abierto

Datos usados en el estudio [DOI 10.15468/dl.0cvdaa](#) [DOI 10.15468/dl.0knvop](#) [DOI 10.15468/dl.0mhjfc](#) [DOI 10.15468/dl.0mrgnu](#) [DOI 10.15468/dl.1dbtku](#) [DOI 10.15468/dl.1ddcv7](#) [DOI 10.15468/dl.1ifkrd](#) [DOI 10.15468/dl.1ip68r](#) [DOI 10.15468/dl.1s9ge0](#) [DOI 10.15468/dl.1ykbdf](#) [DOI 10.15468/dl.21aviy](#) [DOI 10.15468/dl.23club](#) [DOI 10.15468/dl.23f0s0](#) [DOI 10.15468/dl.25hgiq](#) [DOI 10.15468/dl.2cwi9y](#) [DOI 10.15468/dl.2dormk](#) [DOI 10.15468/dl.2efzmx](#) [DOI 10.15468/dl.2fbswj](#) [DOI 10.15468/dl.2kcitl](#) [DOI 10.15468/dl.2poi7m](#) [DOI 10.15468/dl.2qyg4m](#) [DOI 10.15468/dl.2qypfp](#) [DOI 10.15468/dl.2tfnp5](#)

– Si publicar en GBIF (o en otros portales) tenía impacto, ahora se cuanto, y También quien usa “mis datos “ y para qué... y lo puedo demostrar

– Acaba con la “paranoia” de “mis visitas” a “mi web”

<https://www.gbif.org/es/resource/search?contentType=literature&gbifDatasetKey=8354ecee-f762-11e1-a439-00145eb45e9a>

The image shows a screenshot of the GBIF dataset page for the 'Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB)'. The page is in Spanish and includes a header with navigation links, a main title, and a description of the dataset. A callout box labeled '1' points to the dataset description text. A callout box labeled '2' points to the 'Iniciar sesión' button in the top right. A callout box labeled '3' points to the 'DESCARGA' button. A callout box labeled '4' points to the 'Fecha de publicación' and 'Última modificación' fields. A callout box labeled '5' points to the 'Cita' (Citation) section. A callout box labeled '6' points to the 'Commonly used standards' section. A callout box labeled '7' points to the 'DOI' field.

1 Incluye todos los pliegos del herbario, que en la actualidad son cerca de 130.000, en su mayor parte corresponden a plantas de la flora española, con buena representación de la flora europea y americana. Existen ejemplares procedentes de 70 países destacando Portugal, Francia, Finlandia, Chipre, Estados Unidos, Argentina, etc. Como es frecuente, la mayoría de los pliegos es de plantas con semillas, más que de plantas con frutos. Además, en el Herbario se conservan excelentes colecciones de líquenes, 28.000 pliegos, de los cuales hay unos 10.000 briófitos, 15.000 líquenes y 2.500 hongos.

2 Iniciar sesión

3 DESCARGA

4 Fecha de publicación: 8 de enero de 2025
Última modificación de los datos: 8 de enero de 2025
Alojado por: GBIF-Spain
Licencia: CC BY-NC 4.0

5 Cita
MACB (2025). Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB). Version 2.17. Departamento de Biodiversidad, Ecología y Evolución, Unidad Docente de Botánica. Fac. de Ciencias Biológicas, Univ. Complutense de Madrid. Occurrence dataset <https://doi.org/10.15470/akznrb> accessed via GBIF.org on 2025-07-04.

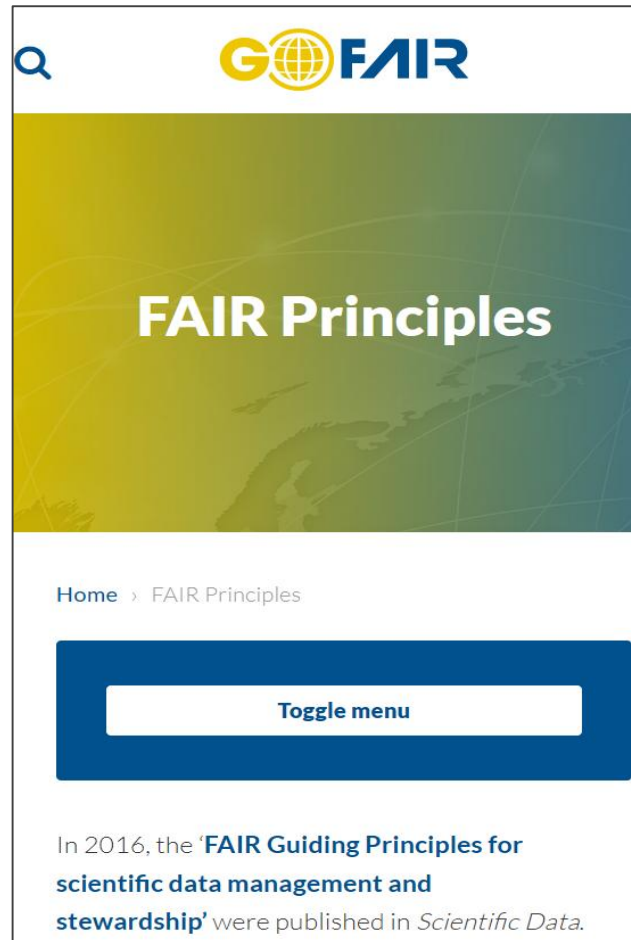
6 Commonly used standards
Darwin Core
The Darwin Core Standard (DwC) offers a stable, straightforward biodiversity data from varied and variable sources. The majority of datasets in GBIF.org rely on 'metadata'—that is, the information that describes the data. Each Darwin Core Archive includes as one of its components an...

7 DOI: 10.15470/akznrb

1. ¿Van los datos con sus metadatos? Quien, donde cuando, como, para qué,...
2. ¿Están indexados?
3. ¿Se pueden descargar?
4. ¿Tienen licencia?
5. ¿Se sabe como citar?
6. ¿Se usa estándares (para los datos y los metadatos)?
7. ¿Están identificados?

<https://www.gbif.org/es/dataset/ceff8c75-950b-4405-a116-60247d1adacb>

Principios FAIR



<https://www.go-fair.org/fair-principles/>

https://red-documentacion.minciencias.gov.co/Gestion_Datos_Investigacion/Principios-FAIR



<https://www.andis.org.au/working-with-data/fairdata/training>

La ultima diapo

- El cambio de paradigma:
 - Del “conocimiento es poder” (y si lo comparto, lo pierdo), a mi relevancia depende de que “lo mío” se utilice
 - El impacto y la relevancia está en que nuestros datos se reutilicen
- Documentar nuestros datos
- En vez de inventar nuestros estándares contribuir a la mejora de los que hay
- Citar las fuentes → DOIs
- Los datos de biodiversidad tienen una larga vida

Francisco Pando

Unidad de Coordinación de GBIF.ES

CSIC

Joaquín Costa 22, 28002 Madrid

pando@gbif.es



<https://creativecommons.org/licenses/by/4.0/deed.es>

GBIF-ES es el Nodo Nacional de Información en Biodiversidad patrocinado por el Ministerio de Ciencia, Innovación y Universidades, gestionado por el Consejo Superior de Investigaciones Científicas (CSIC). <https://www.gbif.es/> <https://datos.gbif.es>, <https://elearning.gbif.es>



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

