

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

16 – 17 de septiembre de 2026, Institut Botànic, Barcelona

Introducción a la publicación de datos de biodiversidad

Francisco Pando
GBIF.ES



**institut
botànic**

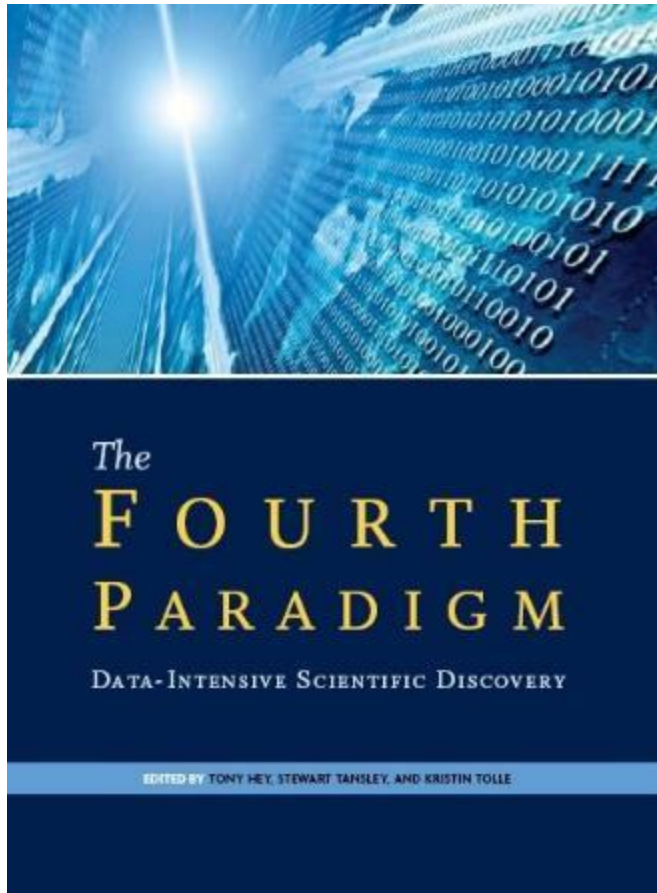
Centre mixt



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*



[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](#)

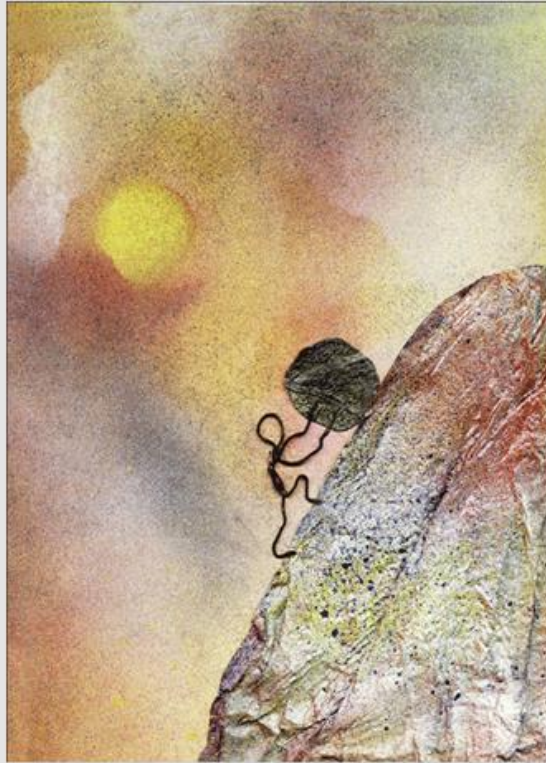
[Communications of the ACM, Volume 67, Issue 12](#) • Page 5 • <https://doi.org/10.1145/3702970>

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 los “productos” de la ciencia (artículos, libros, datos), 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

Hacer los datos abiertos y reutilizables es central en los nuevos paradigmas de la Ciencia

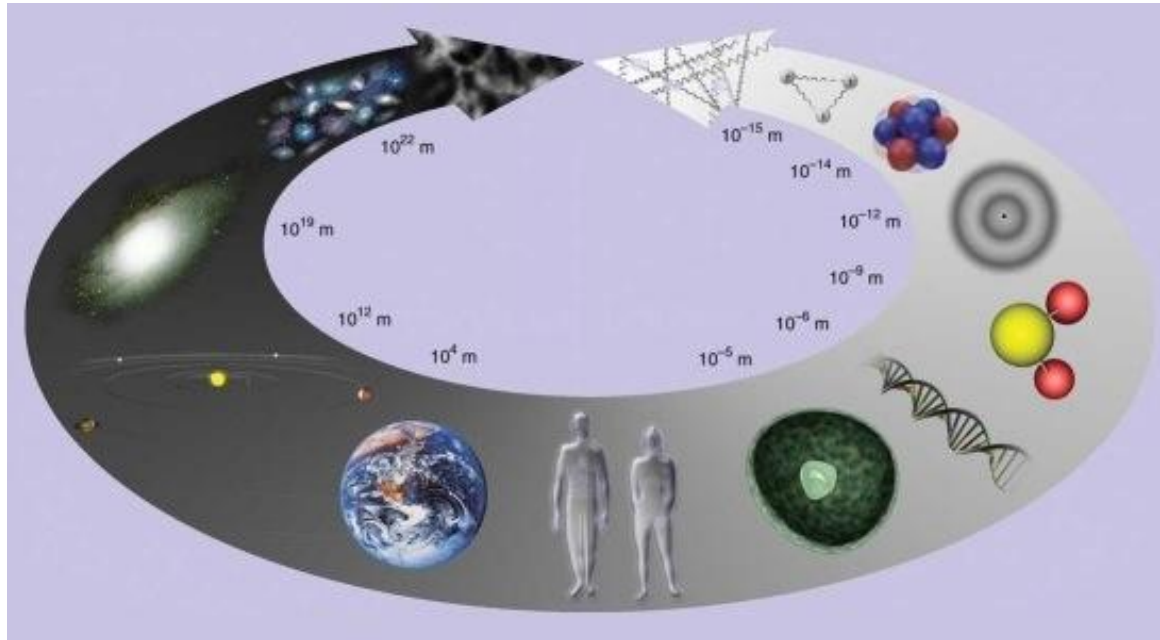


We have oceans of data,
rivers of information, small
puddles of knowledge, and
the odd drop of wisdom

Henry Nix



“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



Retos 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

Free and open access to biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES



3,819,898,830

Occurrence records



123,922

Datasets



2,781

Publishing institutions



15,597

Peer-reviewed papers using

data

Occurrence dataset | Registered 3 April 2007

Col·lecció de plantes de l'herbari de la Universitat de Barcelona

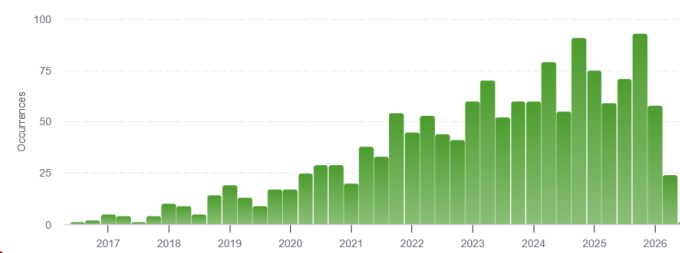
Published by Herbario BCN (CeDocBiV-Univ. de Barcelona)

1,175 citations

124,405 occurrences

Citation metrics

Publication date

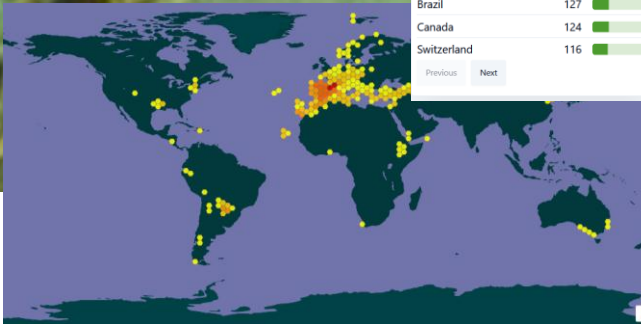


Country or area of researcher
Number of occurrences

129 results

United States of America	424
United Kingdom	245
Germany	230
China	224
Spain	188
France	141
Australia	139
Brazil	127
Canada	124
Switzerland	116

Previous Next



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

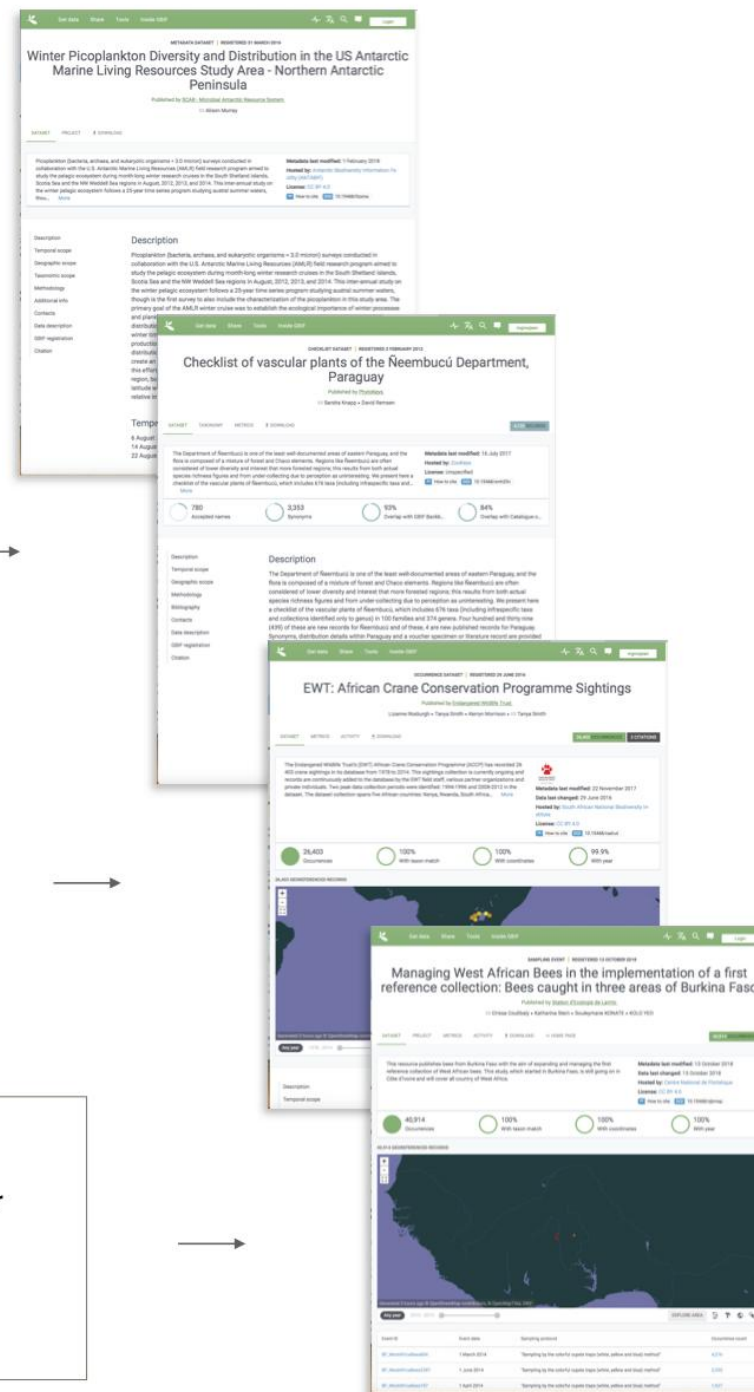
- Scientific names of organisms sharing a common theme or feature (for example: medicinal use).

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.



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

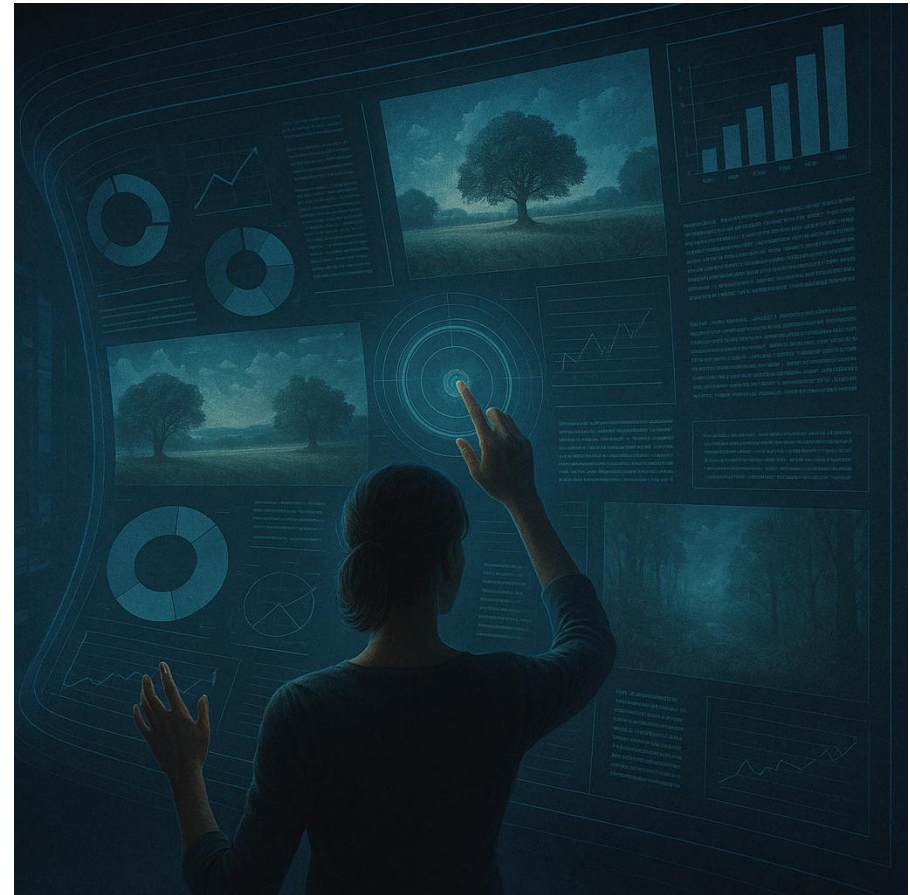


Imagen generada
con IA

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?

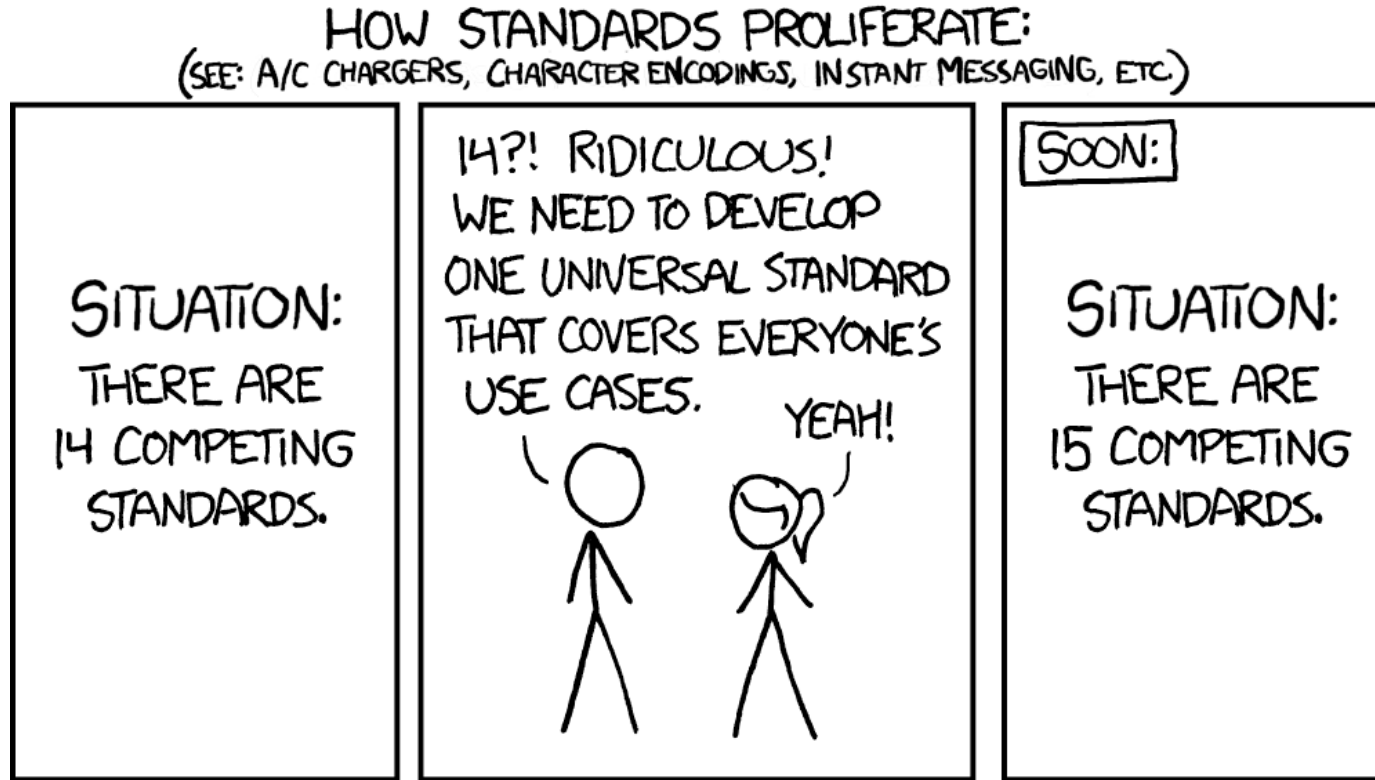
8. Acceso y uso:

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



- 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...”

Biodiversity Information Standards (TDWG)

We are a non-profit organization and a community dedicated to
developing biodiversity information standards.

<https://www.tdwg.org>

ABCD

Access to Biological Collection Data (ABCD) Schema

Audiovisual Core

Audiovisual Core Multimedia Resources Metadata Schema

Darwin Core

Darwin Core

GUID Applicability Statements

GUID and Life Sciences Identifiers Applicability Statements

Latimer Core

Latimer Core, a standard for describing collections

SDS

TDWG Standards Documentation Standard (SDS)

TAPIR

TDWG Access Protocol for Information Retrieval (TAPIR)

TCS

Taxon Concept Schema (TCS)

VMS

Vocabulary Maintenance Standard (VMS)

WGSRPD

World Geographical Scheme for Recording Plant Distributions (WGSRPD)

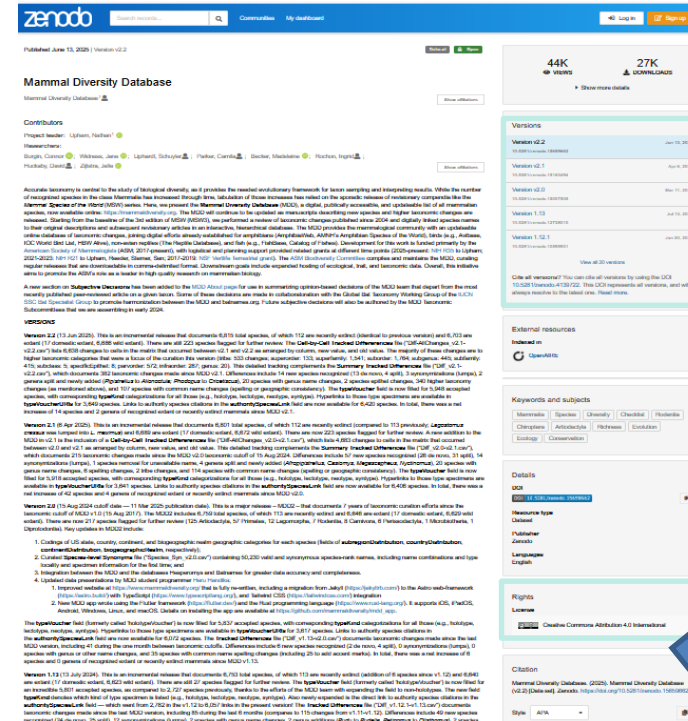
Interest and Task Groups

- MG** Access to Biological Collections Data
- IG** Annotations
- IG** Attribution
 - TG** People in Biodiversity Data
- MG** Audiovisual Core
 - TG** 3D Imagery and Data
 - TG** Views Controlled Vocabularies
- IG** Biodiversity Data Quality
 - TG** Framework on data quality
 - TG** Data quality tests and assertions
 - TG** Data quality use cases
 - TG** Best practices for development of vocabularies of values ("Vocabularies")
- IG** Biodiversity Informatics Curriculum
- IG** Biodiversity Services and Clients
- IG** Biological Interactions Data
- IG** Citizen Science
- IG** Collection Descriptions
 - MG** Latimer Core
 - TG** Minimum Information about a Digital Specimen (MIDS)
 - TG** Modelling Research Expeditions
- MG** Darwin Core
 - TG** Invasive Species
 - TG** Plant Phenology
 - TG** Realm and Biome
 - TG** Sensitive Species Extension
- IG** Earth Sciences and Paleobiology
 - TG** Darwin Core Chronometric Age Extension
 - TG** Extension for Geosciences (EFG)
 - TG** Mineralogy Extension
- IG** Genomic Biodiversity
 - TG** Environmental Samples and eDNA
 - TG** Sustainable DarwinCore MixS Interoperability
- IG** Geographical Schemes
 - TG** Marine Geo Schemes Task Group
 - TG** Terrestrial GeoSchemes TG
- IG** Machine Observations
- IG** Observations & Specimens
 - TG** How Did It Die?
 - TG** Humboldt Extension
 - TG** Material Sample
- IG** Process
- IG** Species Information
 - TG** Invasive Organism Information
 - TG** Plinian Core
- MG** Taxon Concept Schema (TCS) Maintenance Group
 - TG** Taxonomic Concept Schema (TCS2) Task Group
- IG** Technical Architecture Group
 - TG** Standards Mapping Task Group

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

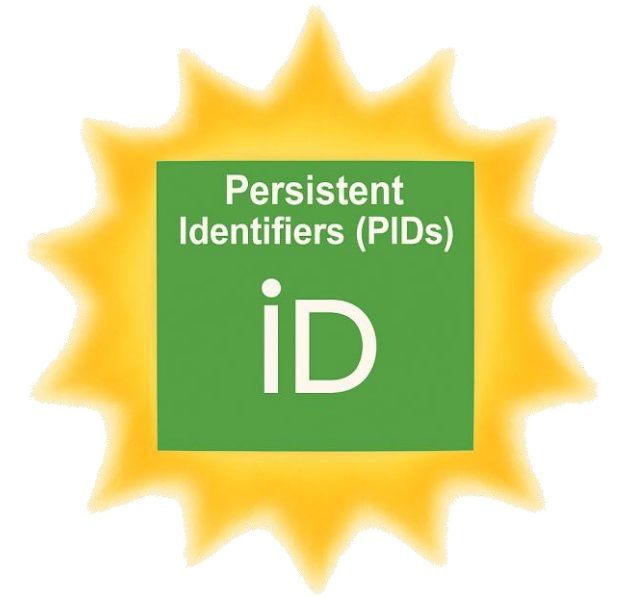
Mammal Diversity Database. (2025). Mammal Diversity Database (v2.2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15659662>

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

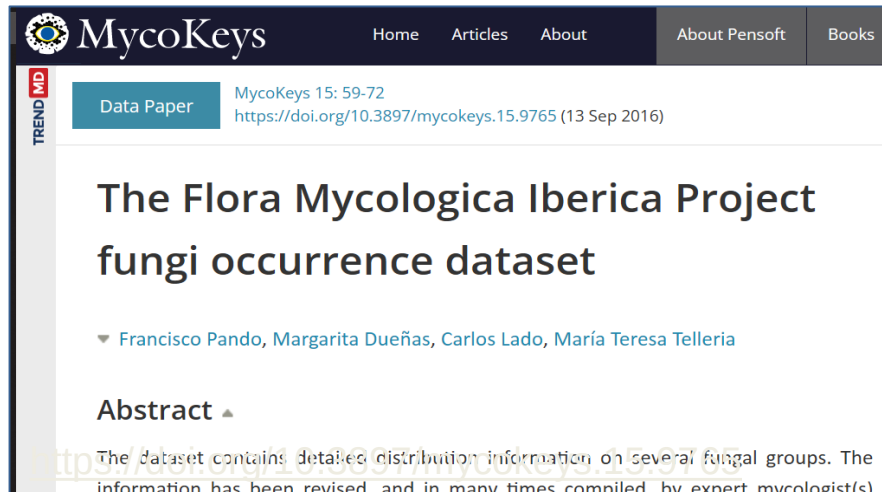


550e8400-e29b-41d4-a716-446655440000

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

Identificadores persistentes, únicos, **resolviles**: DOIs



DOI: Digital Object Identifier

- Ámplia implantación
- Robusto
- Resoluble

trazabilidad

Citación ↔ crédito

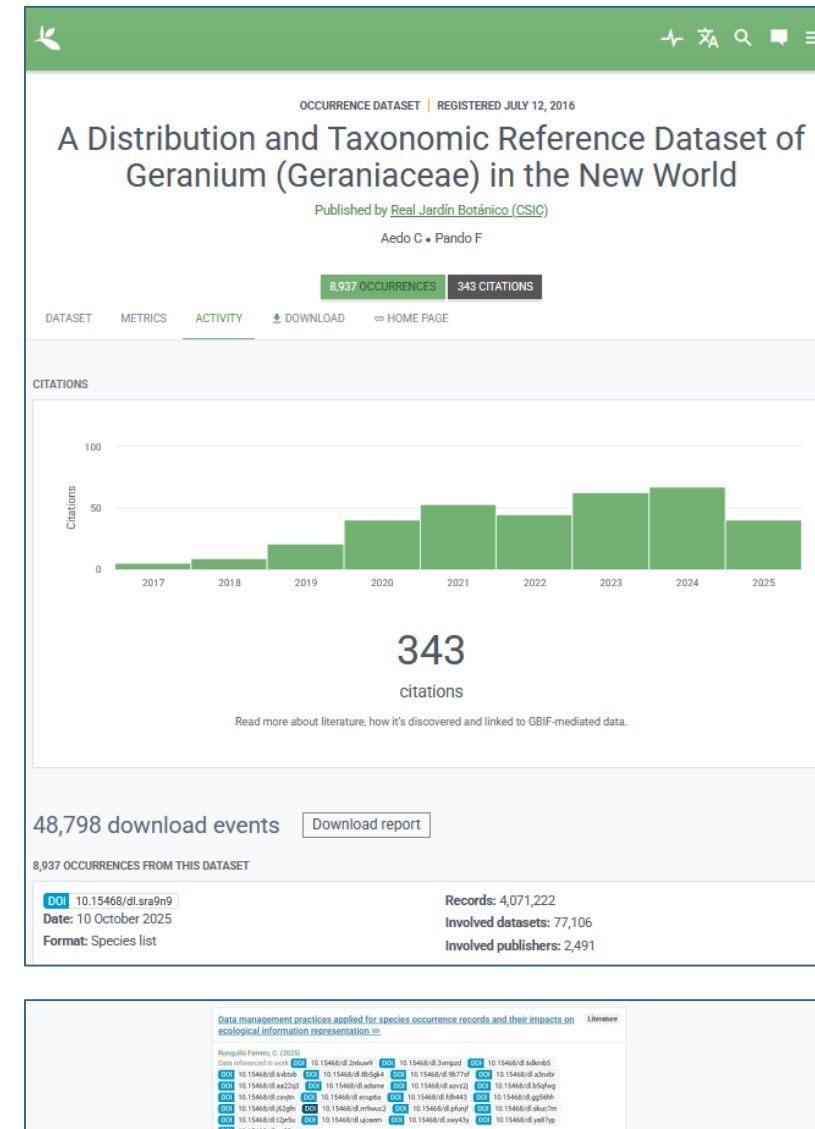
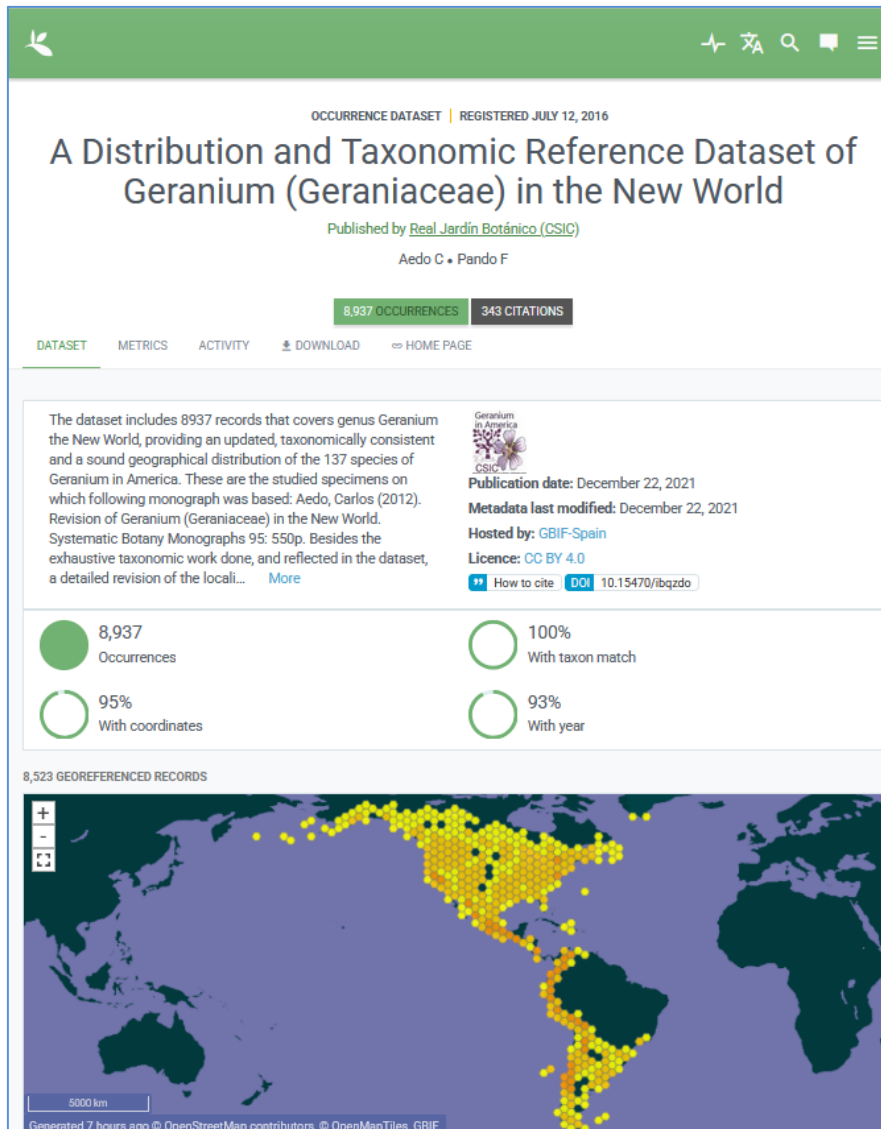


Su uso se amplia más allá de las publicaciones: juegos de datos, gráficos, imágenes, descargas...

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

¿Que cambia los DOIs?

(Identificadores únicos, universales, persistentes y resolvibles)



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

¿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>

El cómo se sirve el dato

The screenshot shows the GBIF dataset page for the 'Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB)'. Numbered callouts highlight the following elements:

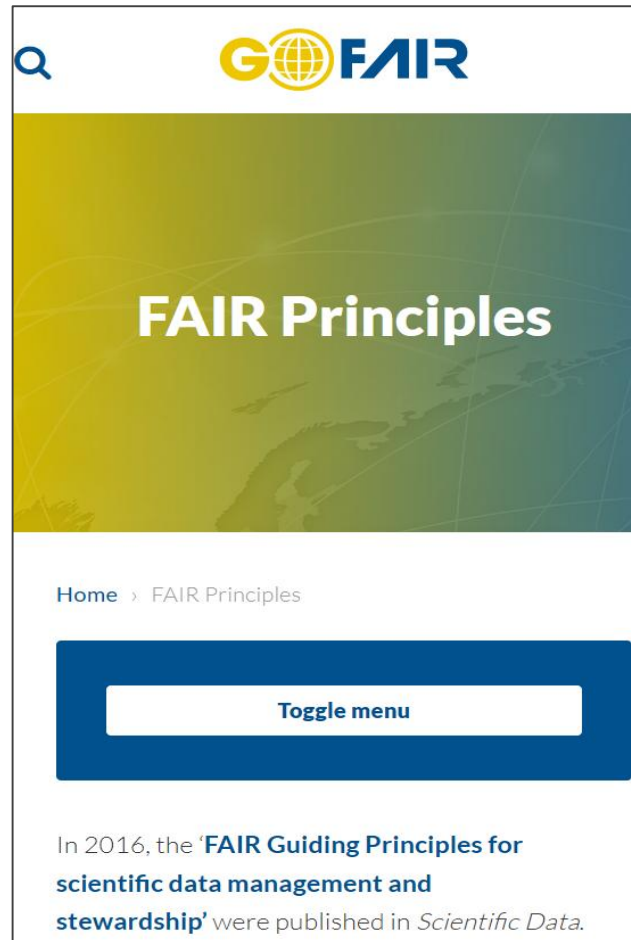
- 1:** A text block describing the dataset's scope, including the number of specimens (130,000) and the types of data (herbarium sheets, seeds, etc.).
- 2:** The 'Iniciar sesión' (Log in) button in the top right corner.
- 3:** The 'DESCARGA' (Download) button.
- 4:** Metadata fields: 'Fecha de publicación: 8 de enero de 2025', 'Última modificación de los datos: 8 de enero de 2025', 'Alojado por: GBIF-Spain', and 'Licencia: CC BY-NC 4.0'.
- 5:** A world map showing the geographic distribution of the dataset, with a 'Cita' (Citation) box overlaid.
- 6:** A sidebar titled 'Commonly used standards' listing 'Darwin Core' and 'EML: Ecological Metadata Language'.
- 7:** The 'Cómo citar' (How to cite) button and the DOI '10.15470/akznrb'.

Additional visible elements include a navigation bar with 'Obtener datos', 'How-to', 'Herramientas', 'Comunidad', and 'Acerca de'; a 'CONJUNTO DE DATOS DE REGISTROS BIOLÓGICOS | REGISTRADO' header; a 'Publicado por' section; a '125.919 REGISTROS BIOLÓGICOS' and '623 CITAS' summary; a '63.511 REGISTROS GEOREFERENCIADOS' map section; and a '2500 REGISTROS CON IMÁGENES' gallery at the bottom.

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
- **Publicar datos es la manera de colocar nuestra colección en el mundo**

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

