

Explorando el ciclo de vida de los datos de biodiversidad y ambientales: desde la recopilación hasta la publicación

Antonio J. Pérez-Luque

Instituto de Ciencias Forestales (CIFOR) | INIA-CSIC (Madrid)

Marzo 04-06, 2025



UNIVERSIDAD
DE GRANADA

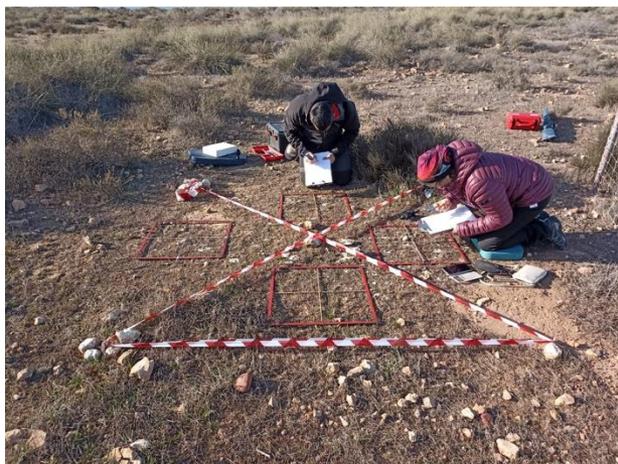
Ciclo de Gestión de los Datos. Ecoinformática
Master Universitario en Conservación, Gestión y Restauración de la Biodiversidad

Importancia de los datos primarios

Antonio J. Pérez-Luque

Instituto de Ciencias Forestales (CIFOR) | INIA-CSIC (Madrid)

2025-03-03



Importancia de los datos primarios

- Ciencia y datos: un binomio indisoluble
- Sobre la importancia de la publicación de datos en diferentes sectores: educativo, empresarial, gubernamental
- Introducción a la publicación de los datos ambientales y de biodiversidad
- ¿Por qué publicar datos científicos?
- Tipos de datos que se pueden publicar
- Repositorios de datos para publicar datos ambientales y de biodiversidad.
- GBIF como ejemplo de publicación de datos de biodiversidad

Ciencia

La ciencia puede concebirse como una **empresa social** cuya finalidad es la **acumulación de conocimiento como bien público** (Munafò et al. 2017).

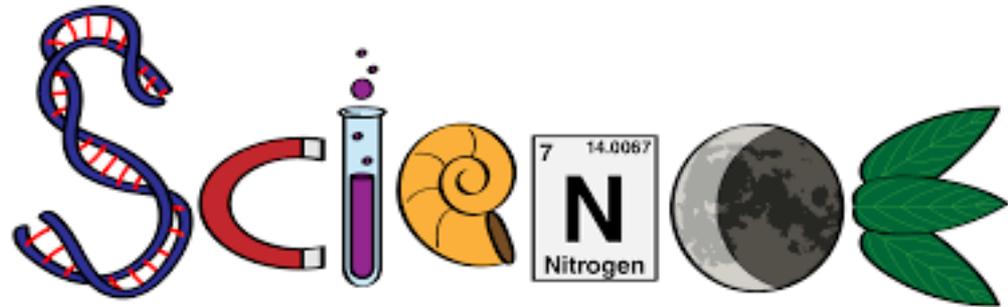


Foto: <https://www.cosesdecinque.blogsek.es/science/>

Derechos Humanos artículo 27

Toda persona tiene derecho a tomar parte libremente en la vida cultural de la comunidad, a gozar de las artes y a participar en el progreso científico y en los beneficios que de él resulten

Ciencia

La ciencia es la empresa organizada y sistemática que reúne conocimientos sobre el mundo y los condensa en leyes y principios comprobables.



Edward O. Wilson

National Geographic

Conjunto de conocimientos obtenidos mediante la **observación** y el razonamiento, sistemáticamente estructurados y de los que se deducen principios y leyes generales (RAE)

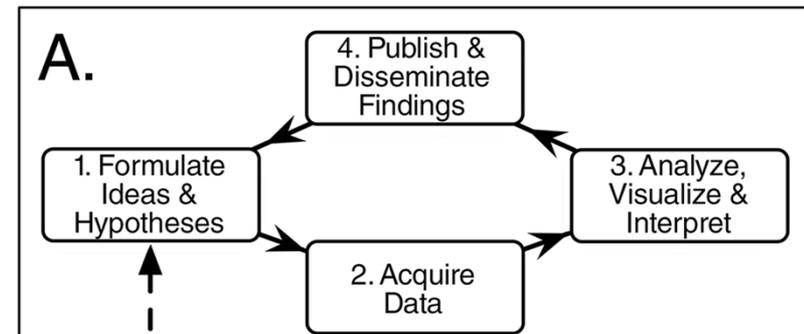
Ciencia & Datos

Uno de los productos de esta gran empresa (la ciencia) es, sin duda, los datos.



Podría incluso afirmarse que se trata de uno de sus productos más valiosos, teniendo en cuenta que constituyen la base para la generación de información y conocimiento científico

Michener 2015



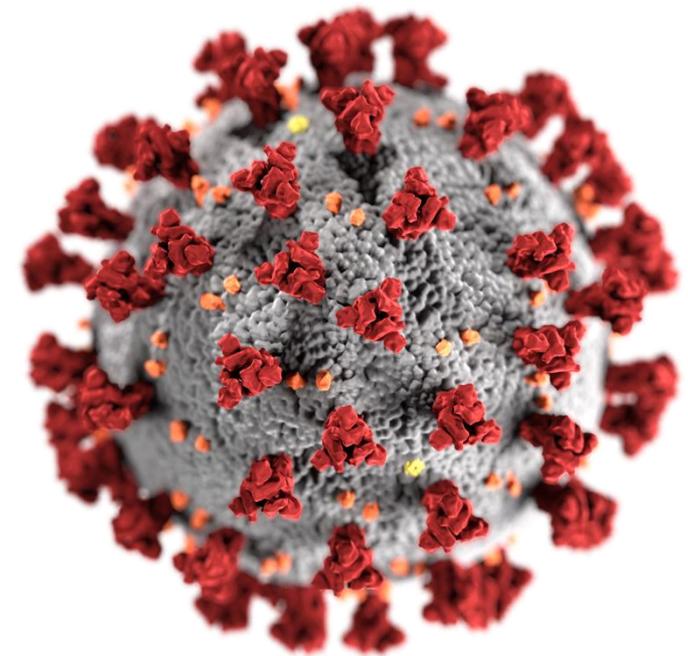
Importancia de los datos

COVID-19

Durante la pandemia de COVID-19, se necesitaban avances médicos urgentes que ofrecieran soluciones a una crisis global. Para ello **se necesitaban datos**.

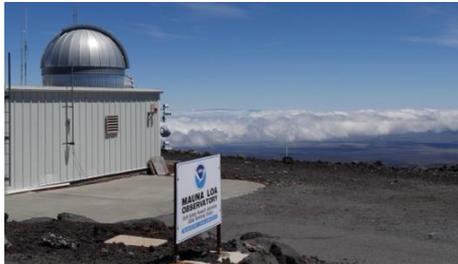
Se necesitaba **compartir datos** de ensayos clínicos, que ayudarán a avanzar en el desarrollo de las vacunas.

Compartir datos puede mejorar la colaboración entre científicos (y con empresas) promoviendo análisis en tiempo real para controlar futuras enfermedades como nuevas variantes del COVID-19



Importancia de los datos

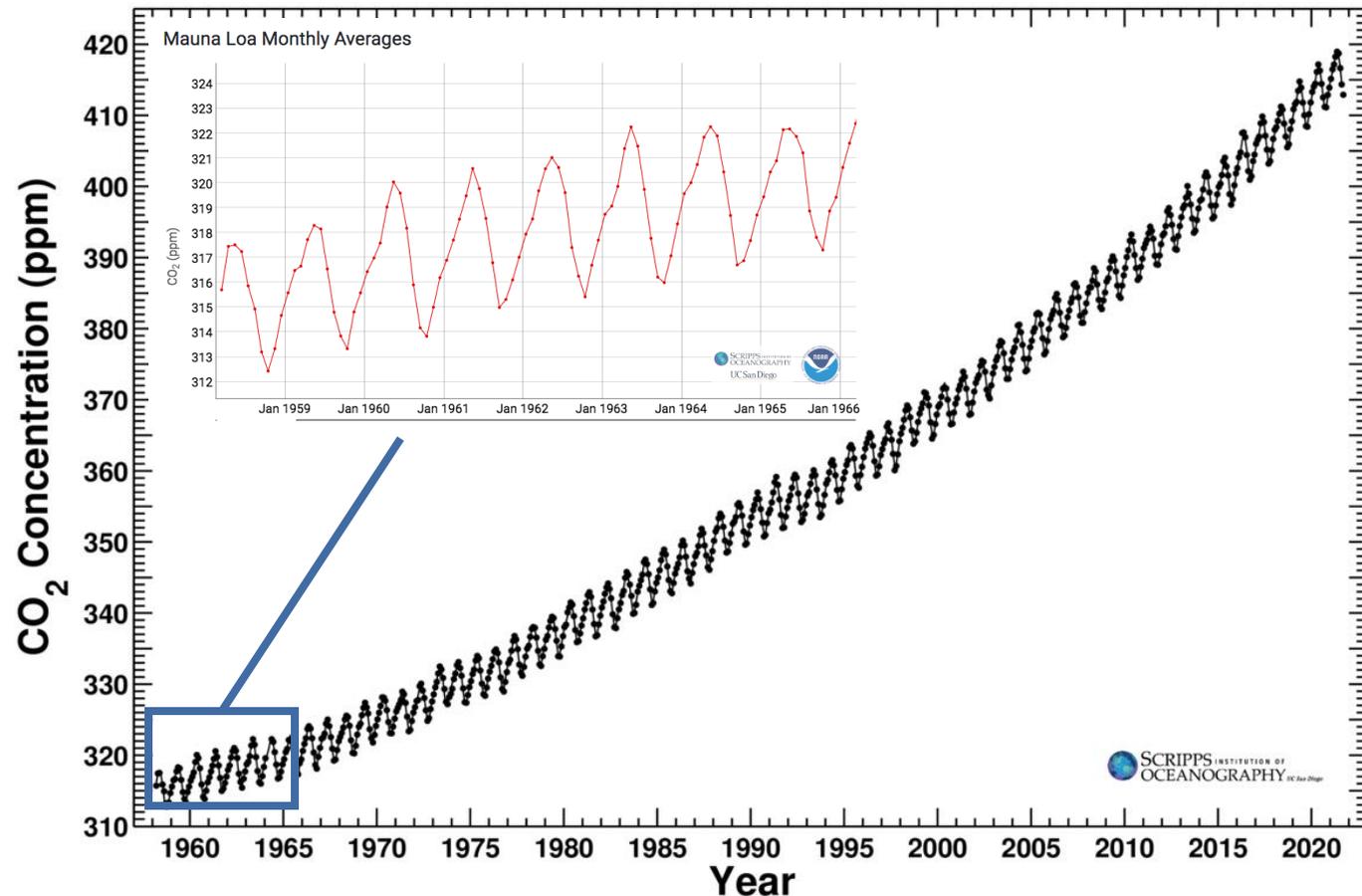
Datos de seguimiento ecológico a largo plazo



Charles Keeling

Mauna Loa Observatory, Hawaii Monthly Average Carbon Dioxide Concentration

Data from Scripps CO₂ Program Last updated October 2021



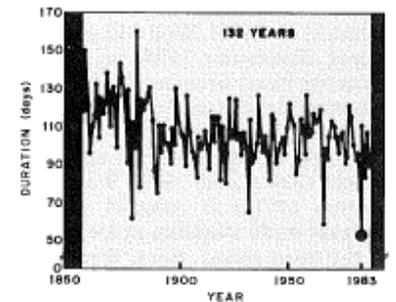
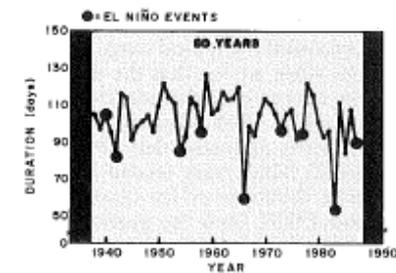
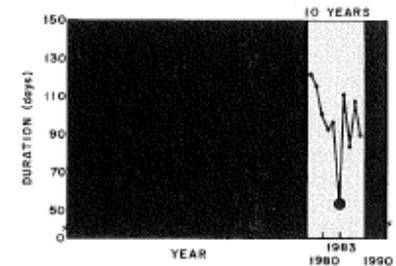
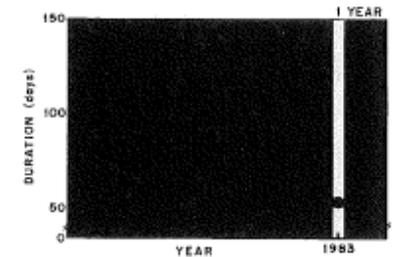
Importancia de los datos

Datos de seguimiento ecológico a largo plazo

Las series de datos a largo plazo permiten probar y desarrollar teorías, así como comprender patrones ecológicos y evolutivos complejos que a menudo surgen en amplias **escalas** temporales y espaciales Kuebbing et al. 2018

La ausencia de series de datos de largo plazo, pueden llegar a provocar graves errores en los intentos de gestionar el medio ambiente

Manguson 1990



Manguson 1990



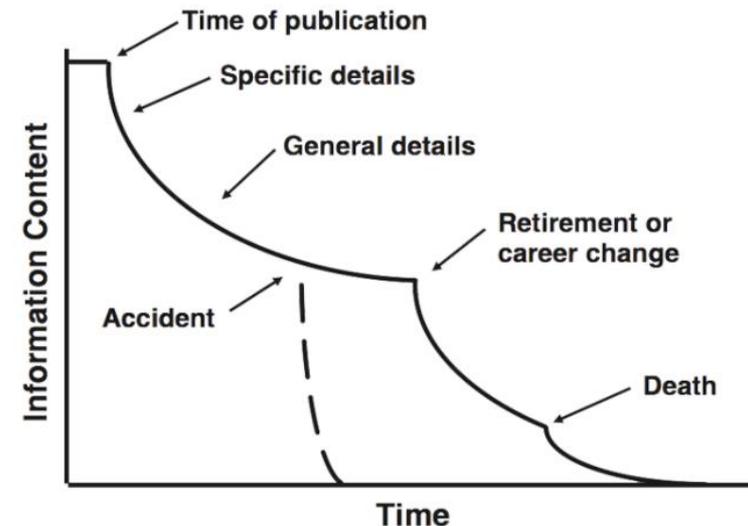
¿Por qué publicar datos?

¿Por qué publicar datos?

La mayoría de los investigadores reconocen la importancia de **compartir datos (beneficios)**, aunque señalan diferentes barreras (**costes**)

Pérez-Luque y Ros-Candeira 2019

- Mejor aprovechamiento de los recursos invertidos en ciencia evitando, por ejemplo, la duplicación de proyectos para obtener datos similares.
- Optimización de metaanálisis y resolución de nuevas cuestiones.
- Reducción del riesgo de pérdida de los datos.



Michener 1995

¿Por qué publicar datos?

- Aumento de la transparencia en ciencia, permitiendo la reproducibilidad de análisis y verificación de resultados
- Incremento en el crédito de los proveedores de datos
- Facilita la accesibilidad a los datos
- Potencia posibles nuevas colaboraciones científicas

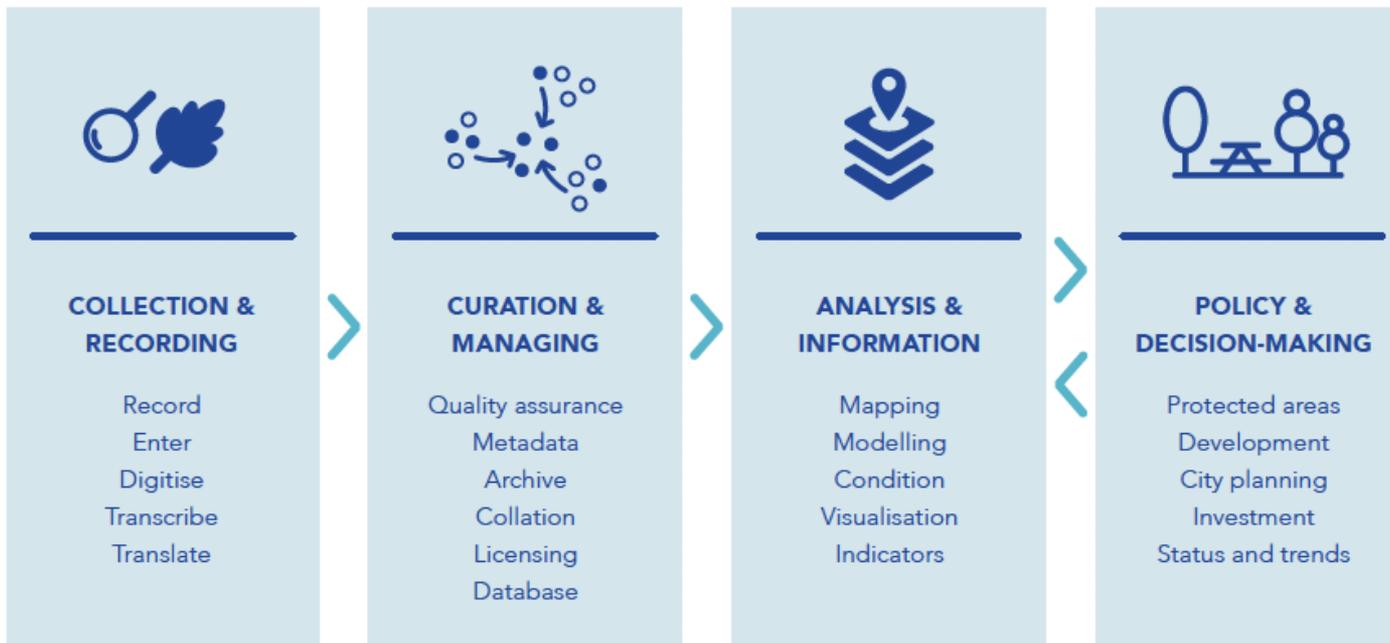
¿Por qué publicar datos?

Decisiones ambientales basadas en conocimiento experto



Better data, better decisions: increasing the impact of biodiversity information

BID-REX - From biodiversity data to decisions: enhancing natural value through improved regional development policies



Ling et al. 2019

Better data, better decisions: increasing the impact of biodiversity information

<https://www.interregeurope.eu/bid-rex/library/#folder=1000>

¿Por qué publicar datos?

Who are the main users of biodiversity data?

Local, national, regional and global scale biodiversity data are needed by:

National Governments

Policy reporting, spatial planning, natural capital accounting, biodiversity screening (e.g. to issue licenses for concessions), ...

Inter- & Non-Governmental Organisations

Policy development & implementation, assessments (e.g. IPBES), indicators & targets (e.g. Aichi Targets, Sustainable Development Goals), spatial conservation planning, ...

Corporate sector

Screening (e.g. when bidding for concessions), Environmental Impact Assessment, ...

Research bodies

Mapping, species distribution modelling, blue carbon assessment, ecosystem service valuation, protected area design...

¿Por qué publicar datos?

Biological Conservation 213 (2017) 335–340



Unlocking the flow of biodiversity data for decision-making in Africa



P.J. Stephenson ^{a,*}, Nadine Bowles-Newark ^b, Eugenie Regan ^{b,2}, Damon Stanwell-Smith ^{b,3}, Mallé Diagana ^{c,4}, Robert Höft ^d, Harouna Abarchi ^e, Tanya Abrahamse ^f, Christine Akello ^g, Hilary Allison ^b, Olaf Banki ^m, Barthelemy Batieno ^h, Samuel Dieme ⁱ, Arthur Domingos ^j, Russell Galt ^f, Cicilia W. Githaiga ^k, Abdoulaye Bine Guindo ^l, David L.N. Hafashimana ^g, Tim Hirsch ^m, Donald Hobern ^m, John Kaaya ⁿ, Ronald Kaggwa ^g, Martha Mphatso Kalemba ^o, Ibrahim Linjouom ^p, Budu Manaka ^f, Zawadi Mbwambo ^q, Monipher Musasa ^o, Eric Okoree ^r, Aggrey Rwetsiba ^g, Ahmat Brahim Siam ^s, Adjima Thiombiano ^t

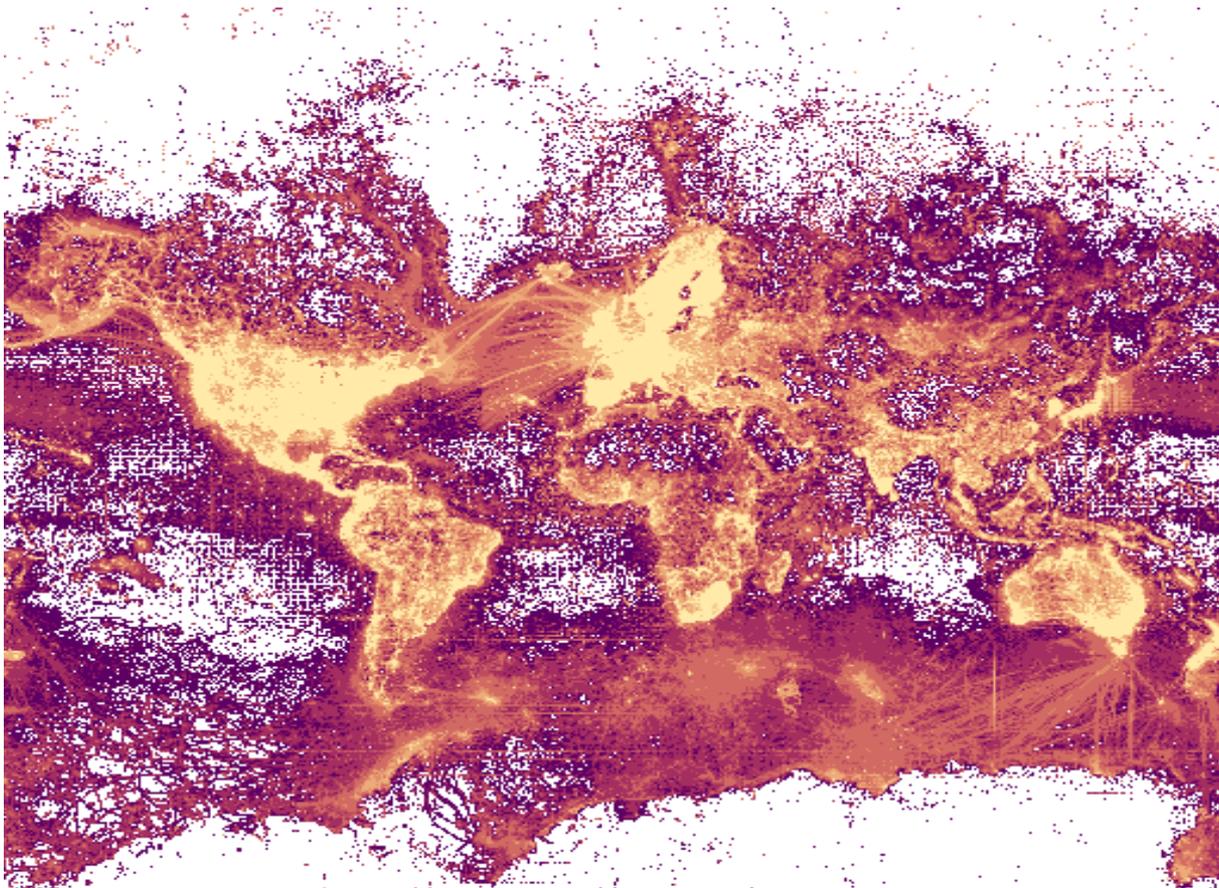
Stephenson et al. 2017

<https://doi.org/10.1016/j.biocon.2016.09.003>

- Muchas agencias gubernamentales no pueden acceder o utilizar los datos sobre biodiversidad que necesitan para tomar decisiones para la gestión ambiental y económica.
- Necesidad de datos, que sean de **calidad** y que se puedan utilizar

¿Cómo lo estamos haciendo?

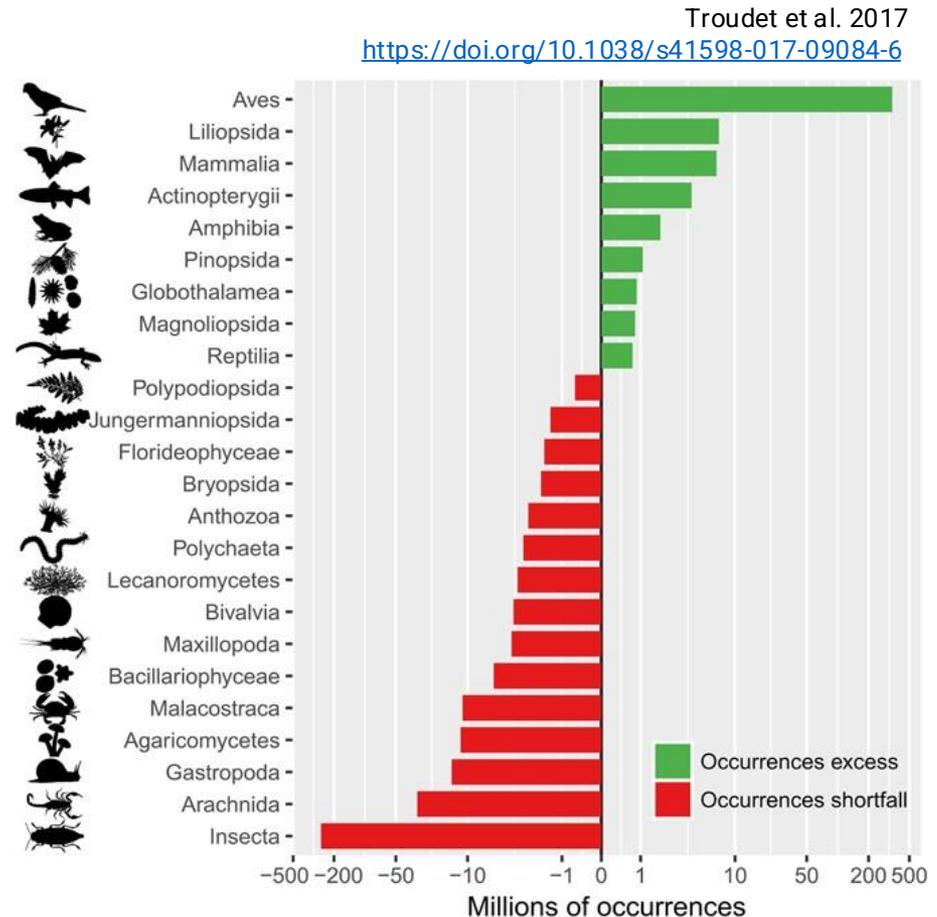
Representación espacial de los datos de ocurrencias integrados en GBIF



¿Cómo lo estamos haciendo?

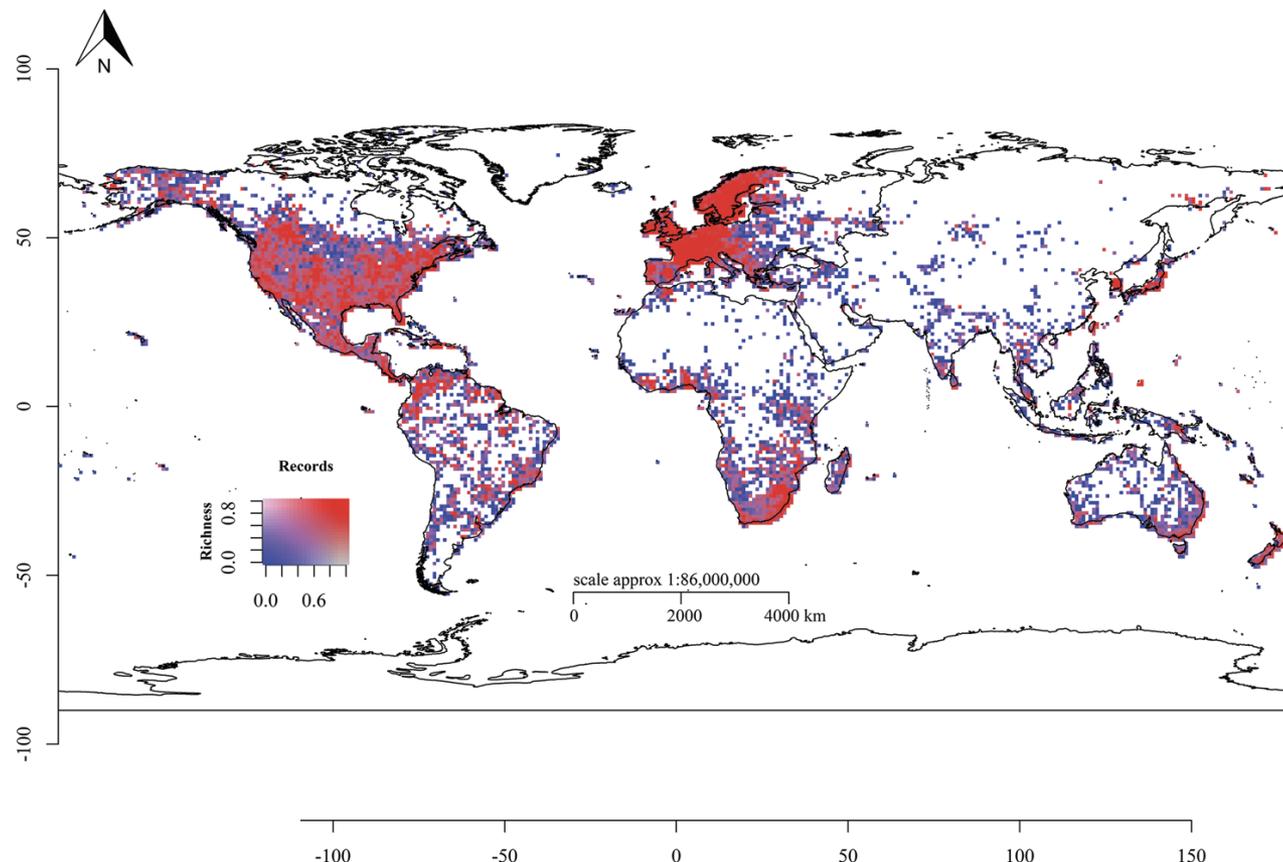
Sesgos taxonómicos

Existen mas datos de
ocurrencias de unos grupos
taxonómicos que de otros



¿Cómo lo estamos haciendo?

Sesgos espaciales



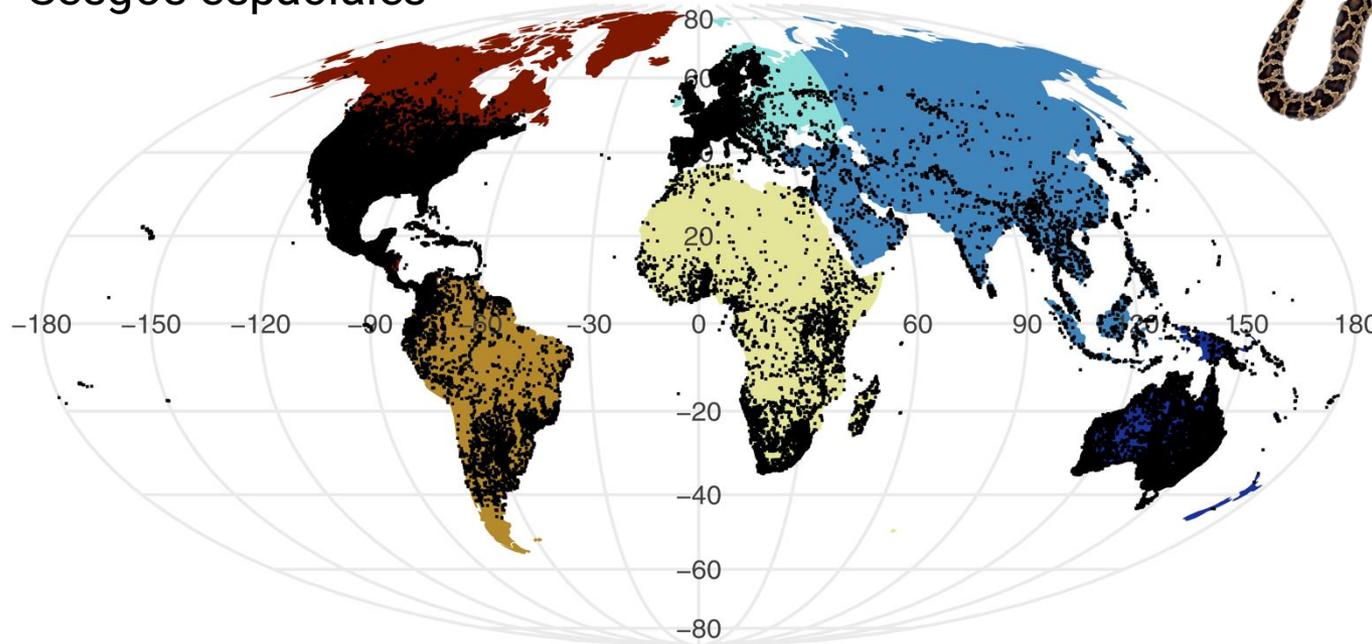
Ecological Entomology (2021), DOI: 10.1111/een.13027

Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction

MAYA ROCHA-ORTEGA,¹ PILAR RODRIGUEZ²
and ALEX CÓRDOBA-AGUILAR¹ ¹Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico and ²Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, CONABIO, Delegación Tlalpan, Mexico

¿Cómo lo estamos haciendo?

Sesgos espaciales

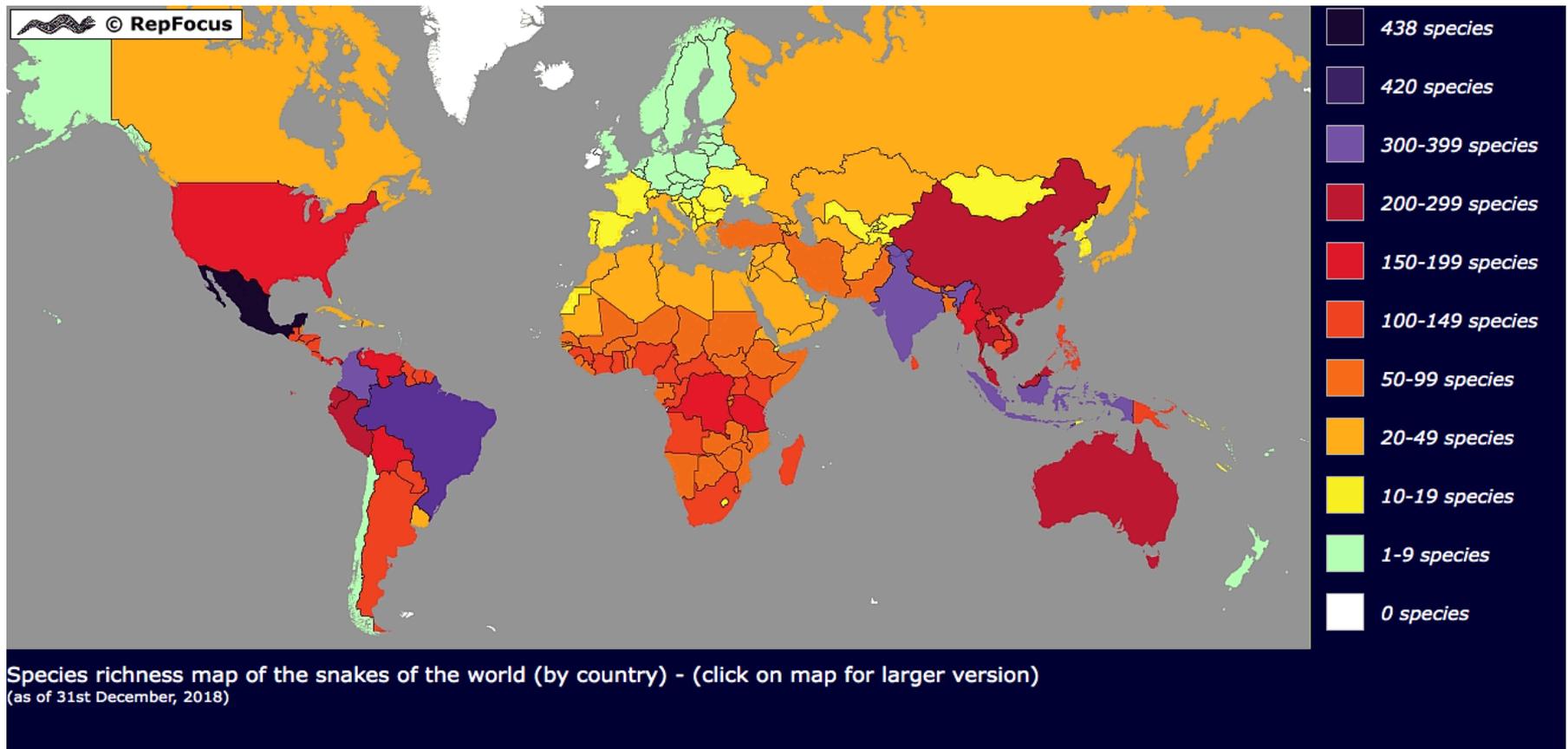


82 % de los registros
Norte América, Europa
y Ocenania

Grandes huecos en
áreas tropicales (zonas
con mayor diversidad)

Marshall & Strine 2019

¿Cómo lo estamos haciendo?



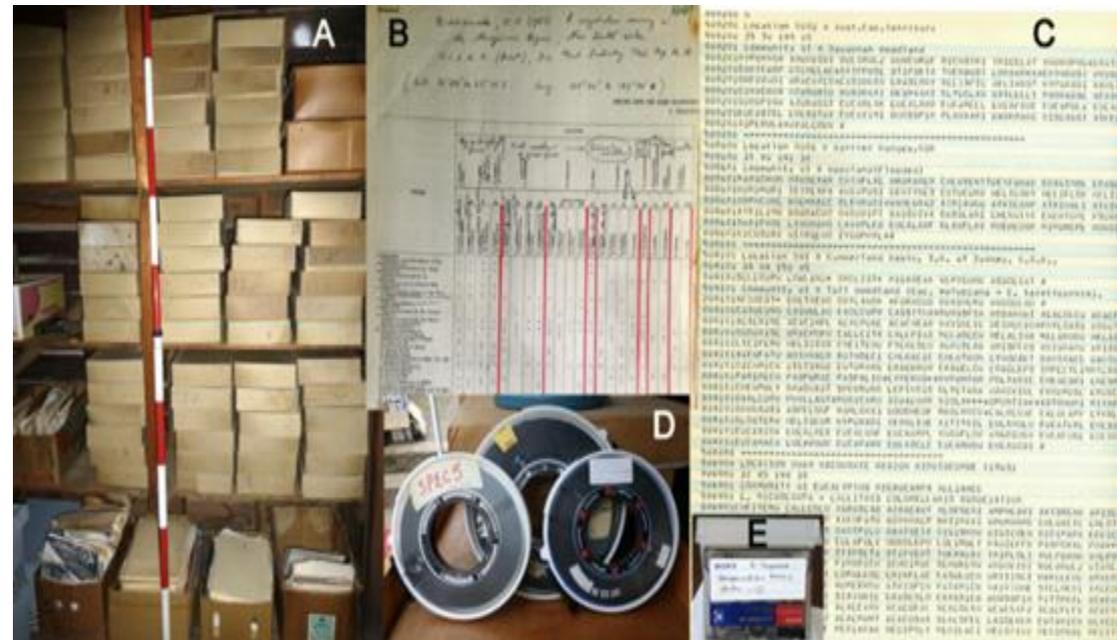
¿Cómo lo estamos haciendo?

Roche et. al. 2015 <https://doi.org/10.1371/journal.pbio.1002295>

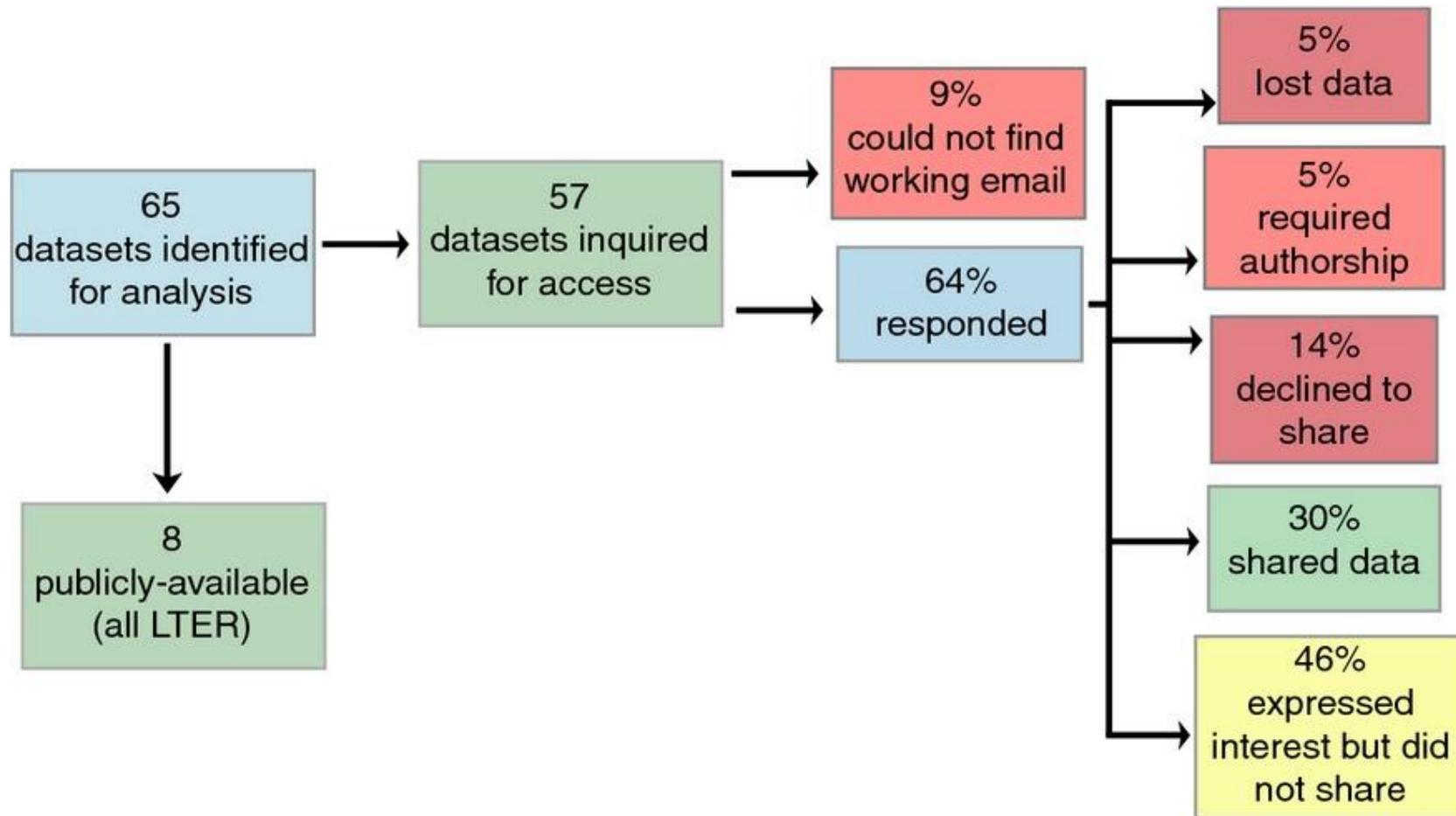
n = 100 conjuntos de datos en revistas de Ecología y Evolución.

56% estaban incompletos. De los archivados, el 64% estaban archivados de manera que parcial o totalmente impedían la reutilización.

Datos en “**peligro de extinción**”



¿Cómo lo estamos haciendo?



Beneficios directos (“para ti, para mi”) de publicar datos ... lo antes posible

Ordenar los datos lo antes posible

Publicar datos promueve una mayor atención a que los datos estén mas ordenados y presentables

Datos mas legibles

Los datos incluidos en un repositorio pueden estar accesibles de forma remota y leerse iterativamente (e.g. script R – GBIF) para realizar control de calidad

Simplifica el versionado y el provenance

Los repositorios permiten el versionado de los datos, lo cual asegura una ciencia reproducible y transparente. También reduce la confusión.

Los datos están archivados y en línea

Publicar los datos en repositorios es un seguro contra la pérdida de datos

Beneficios directos (para ti, para mi) de publicar datos ... lo antes posible

Los metadatos son descritos de forma mas clara y facilita la escritura de informes, artículos, etc.

La redacción de metadatos desarrolla un pensamiento claro y permite una descripción eficaz de los datos. Además facilita la redacción de los métodos y resultados para un informe / publicación

Comprobación de errores

Los colectores de datos, los autores, así como otras personas ajenas al equipo que ha tomado los datos, pueden detectar errores mas fácilmente

Se facilita el intercambio de datos

Tanto dentro del equipo de trabajo como fuera del mismo. Frente a llenar el correo de versiones de archivos de datos

Da crédito a los datos

La publicación de artículos es lenta, y publicar los datos ayuda a dar crédito y establece un objeto citable.

Aumenta la probabilidad de que un trabajo sea aceptado

Los datos abiertos facilitan la revisión por pares, y demuestra un compromiso por la ciencia abierta. Además los autores pueden citar el conjunto de datos en el artículo.



**¿Obstáculos, costes,
inconvenientes para
publicar datos?**



Tipos de datos que se pueden publicar

Datos de colecciones naturales

Varios trabajos han demostrado su potencial para comprender procesos ecológicos en la era del cambio global

Meineke et al. 2018

<http://dx.doi.org/10.1098/rstb.2017.0386>

Bartomeus et al. 2018

<http://dx.doi.org/10.1098/rstb.2017.0389>

1.2 – 2.1 x 10⁹ especímenes en colecciones naturales (> 90% sin digitalizar)

Peterson et al. 2018

<https://doi.org/10.3897/BDJ.6.e26826>



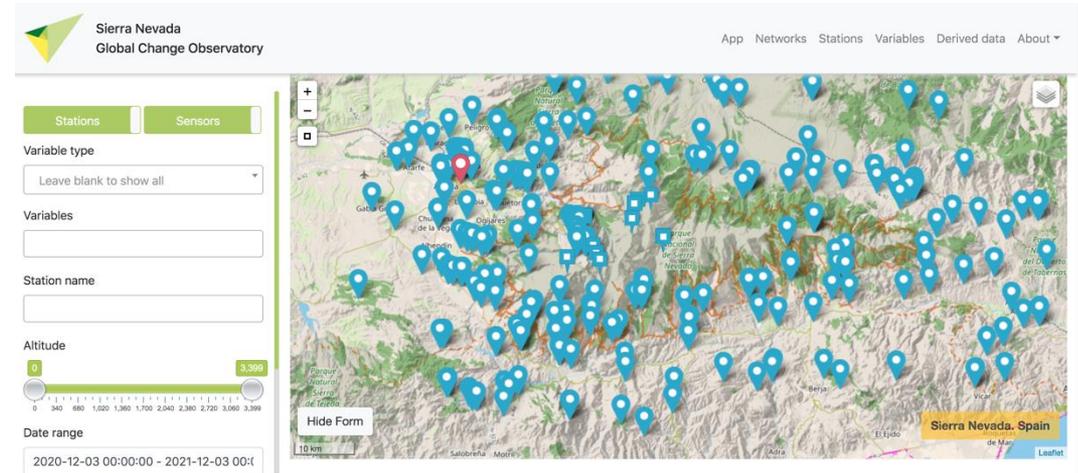
Funk 2018

<https://doi.org/10.1111/jse.12315>

Datos climáticos

Climanevada

- Estaciones meteorológicas
- Sensores climáticos
- Datos de tesis doctorales, proyectos de investigación, etc.



Pérez-Luque et al. 2021

<https://doi.org/10.7818/ECOS.2155>



Imágenes



Imágenes

Tour Flandes (Bélgica)

> 200 h TV

1981 – 2016

46 especies arbóreas

67 % Aumento de floración

(a)



(b)



(c)



De Frenne et al. 2018

<https://doi.org/10.1111/2041-210X.13024>

Datos de seguimiento ambiental / ecológico

Project name Home Temporal coverage Geographic coverage Data tables Custom units Other entities

Hourly soil temperature data from Juniper-thicket ecosystems of Sierra Nevada LTER (Spain)

Data table(s) 2

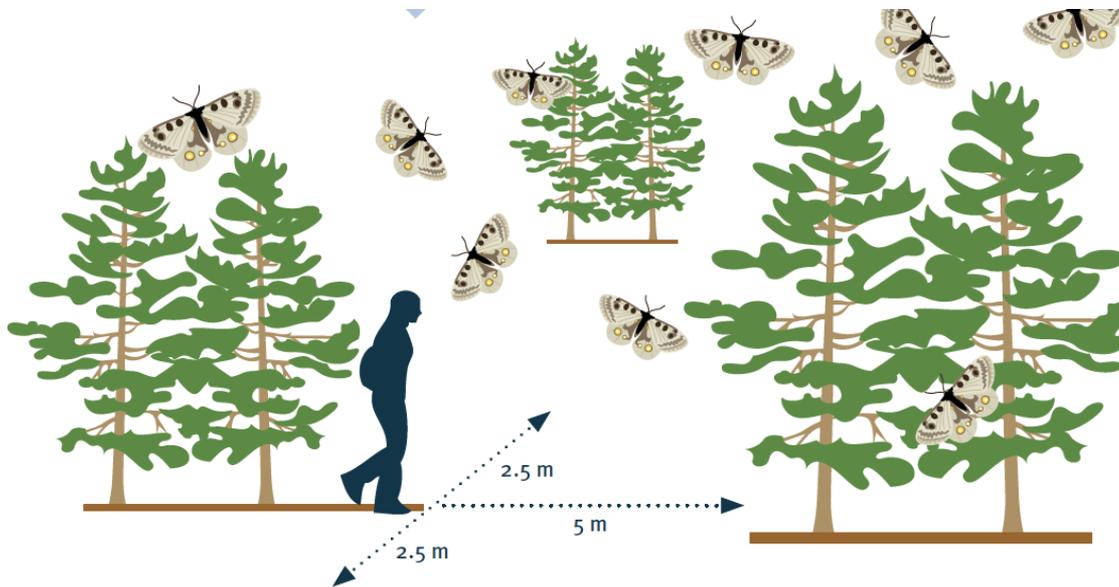
Other entity(ies) 0

Abstract

Hourly temperature data from the ADAPTAMED Juniper-thicket plots located at two contrasting slopes (northern and southern) of Sierra Nevada mountain (Spain). Since 2017 several temperature loggers (n=21) were set up at five different microhabitat within High-mountain scrublands: stone, shrubland (*Genista* species), shrubland (*Juniperus communis*), open ground, and wet meadow. The sensors (HOBO TidbiT Temperature Data Logger) were placed at 5 cm depth. They record data every hour. Data were cleaned and validated by range.



Transectos de estimación da abundancia de mariposas, aves



EVENUTO DE MUESTREO | REGISTRADO

Dataset of butterfly monitoring in Sierra Nevada (Spain)

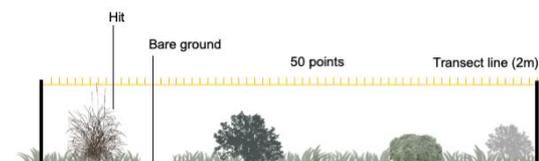
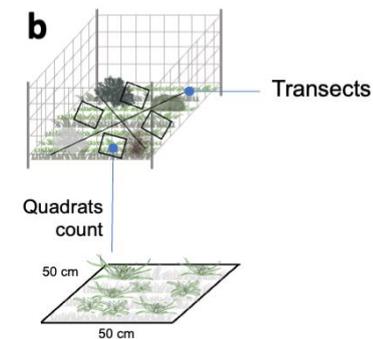
Publicado por Sierra Nevada Global Change Observatory, Andalusian Environmental Center, University of Granada, Regional Government of Andalusia

Barea-Azcón J M

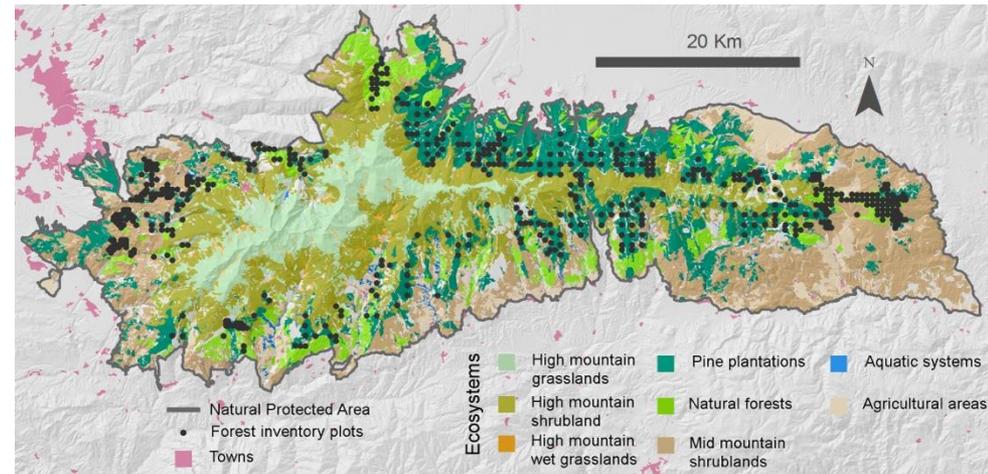
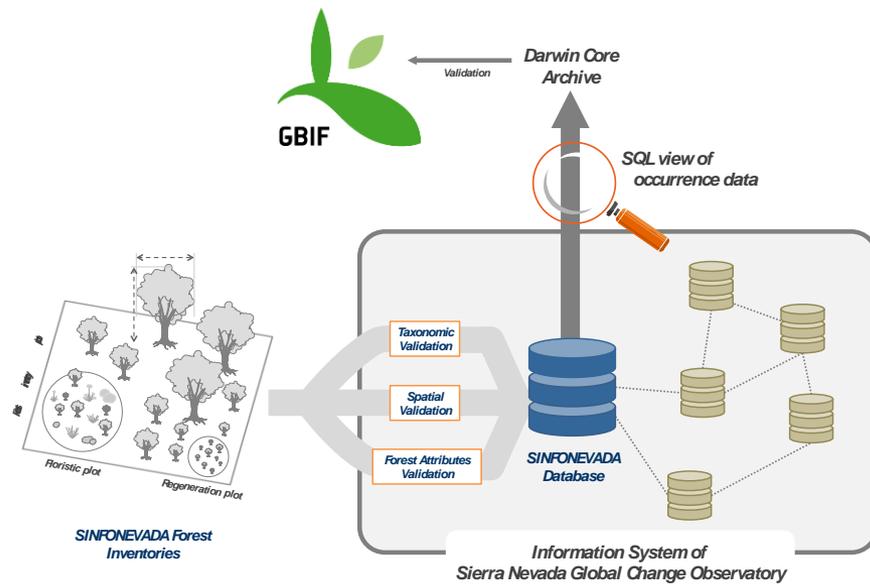
<https://www.gbif.org/es/dataset/d7a3df60-e856-438b-851d-74653ef6c0d2>



Muestras clásicas de vegetación



Inventarios Forestales



Muestreos Específicos asociados a una investigación

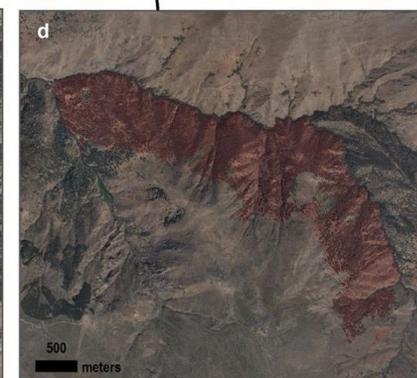
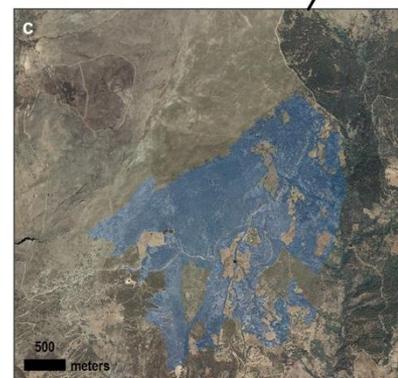
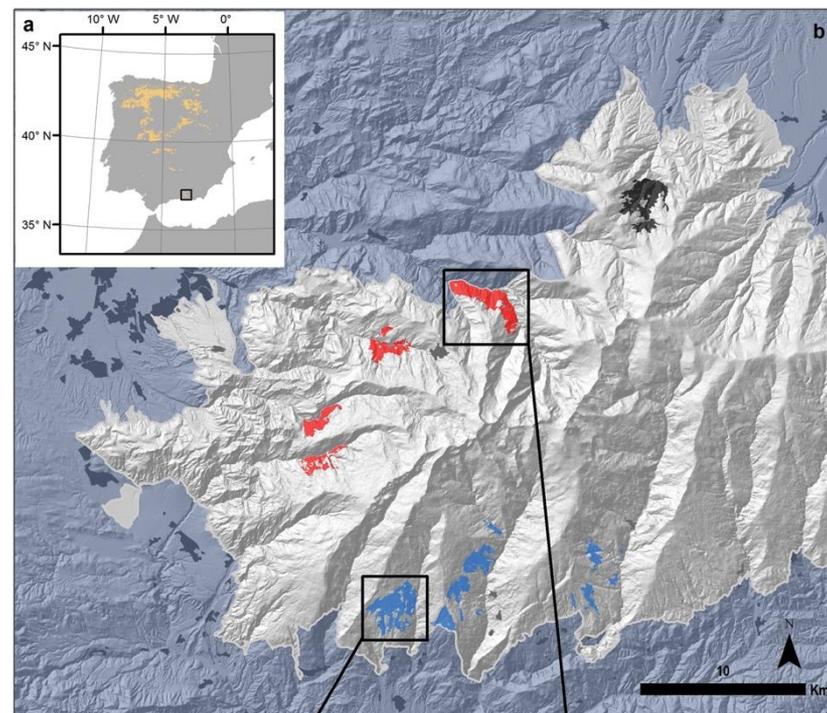
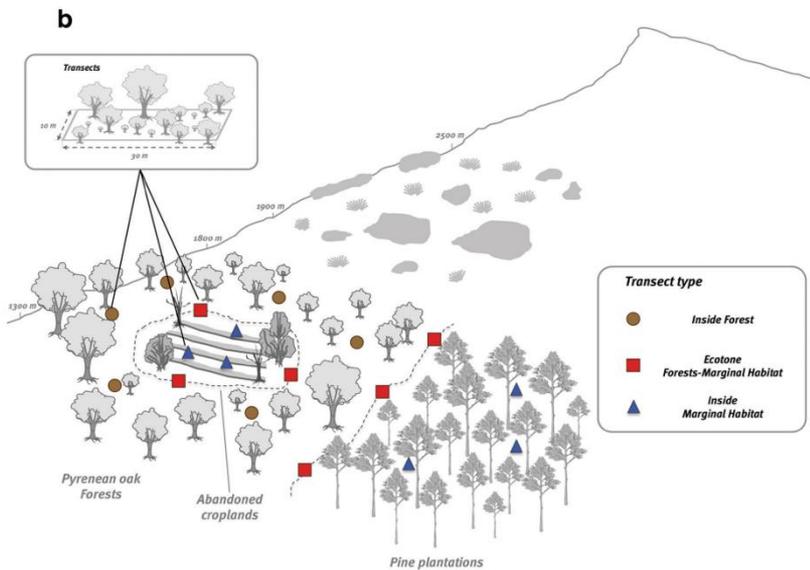
PhytoKeys 56: 61–81 (2015)
doi: 10.3897/phytokeys.56.5482
<http://phytokeys.pensoft.net>

DATA PAPER

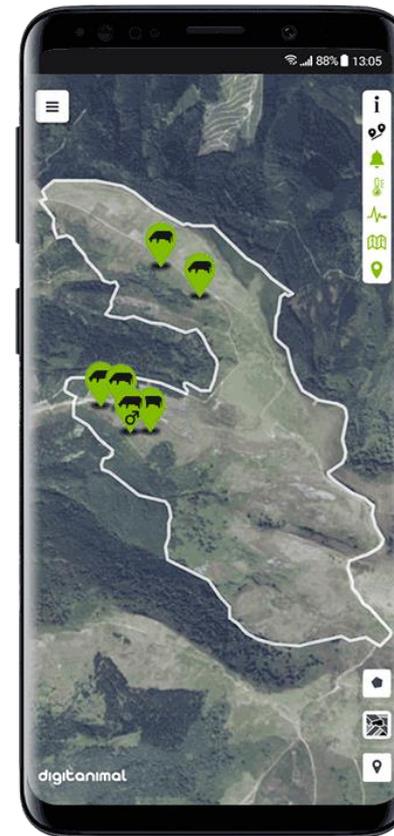
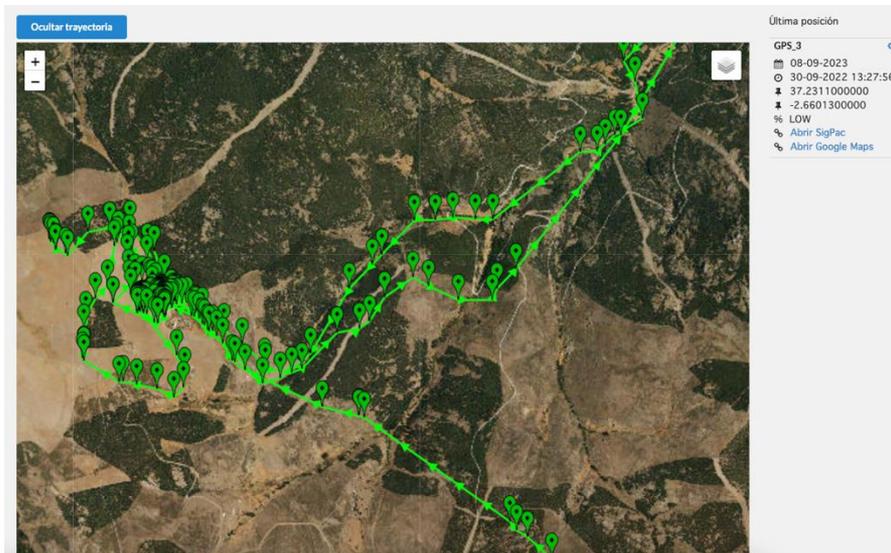
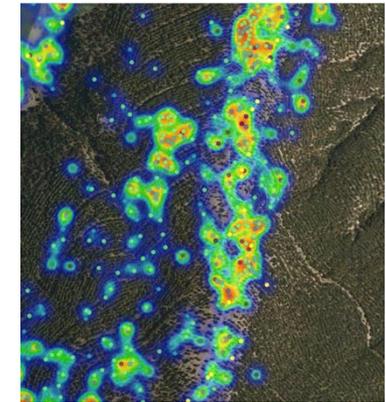
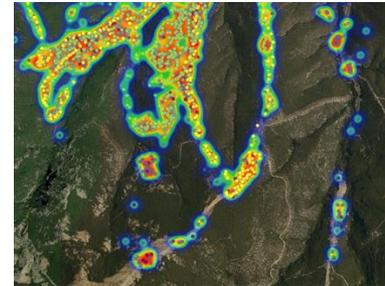
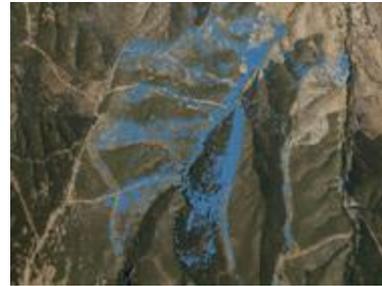
PhytoKeys
A peer-reviewed open-access journal
Launched to accelerate biodiversity research

Dataset of MIGRAME Project (Global Change, Altitudinal Range Shift and Colonization of Degraded Habitats in Mediterranean Mountains)

Antonio Jesús Pérez-Luque^{1,2}, Regino Zamora^{1,2},
Francisco Javier Bonet^{1,2}, Ramón Pérez-Pérez^{1,2}



Datos de movimiento



Datos de movimiento

Virtual Research Environment Grazing Intensity

Pérez-Luque et al. 2023. [10.5281/zenodo.7347993](https://doi.org/10.5281/zenodo.7347993)

vreGI



vre Grazing Intensity

Data selection

1. Select Protected Area
Sierra Nevada

2. Select livestock farmer
SNE-4

3. Filter by date
2022-01-25 to 2023-07-31

Plotting options

Heatmap

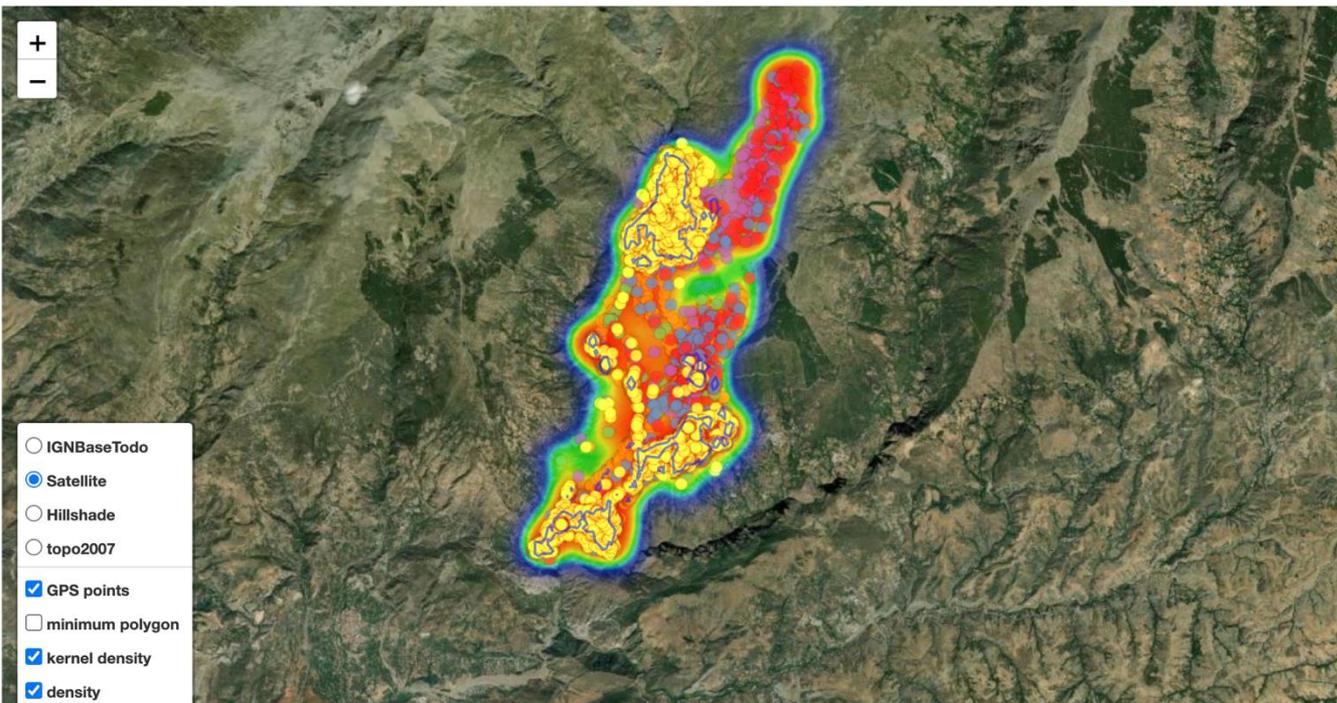
Minimum convex polygon

Kernel Density (95%)

Download shapefiles (.shp)

View Data in Map

Map Table More info



IGNBaseTodo

Satellite

Hillshade

topo2007

GPS points

minimum polygon

kernel density

density

Leaflet | © Instituto Geográfico Nacional de España, Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community



Repositorios de datos

Repositorios de datos

Generales



Buscadores de Repositorios

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

<https://www.nature.com/sdata/policies/repositories>



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search

Repositorios de datos

Temáticos



ISRIC
World Soil Information

WoSIS Soil Profile Database

The World Soil Information Service (WoSIS) aims to serve the user with a selection of standardised and ultimately harmonised soil profile data. The quality-assessed data may be used to underpin digital soil mapping and a range of global assessments. WoSIS is an important building block of ISRIC's evolving, searchable data infrastructure.



<https://www.isric.org/explore/wosis>

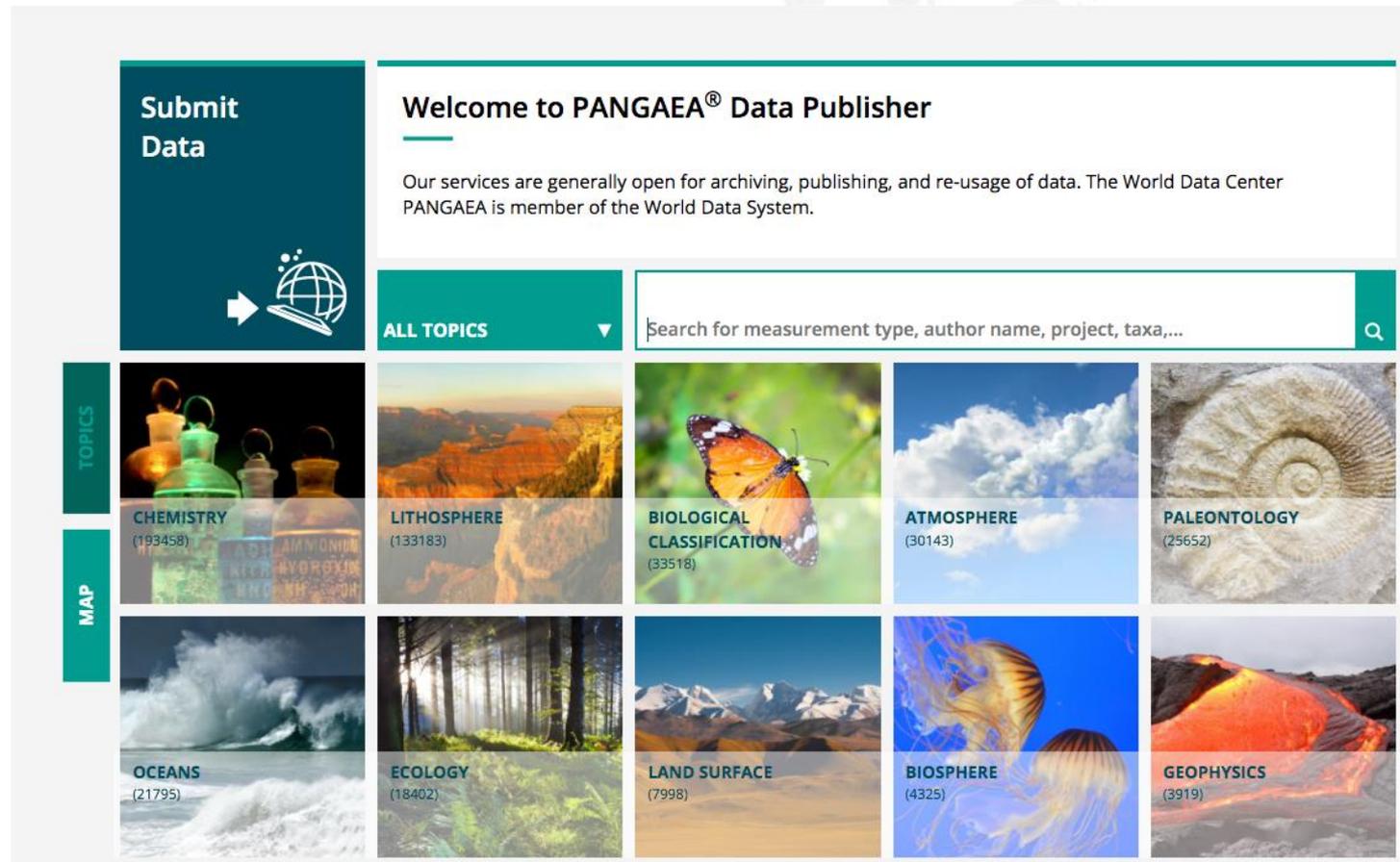
Repositorios de datos

Temáticos



PANGAEA.

Data Publisher for Earth & Environmental Science



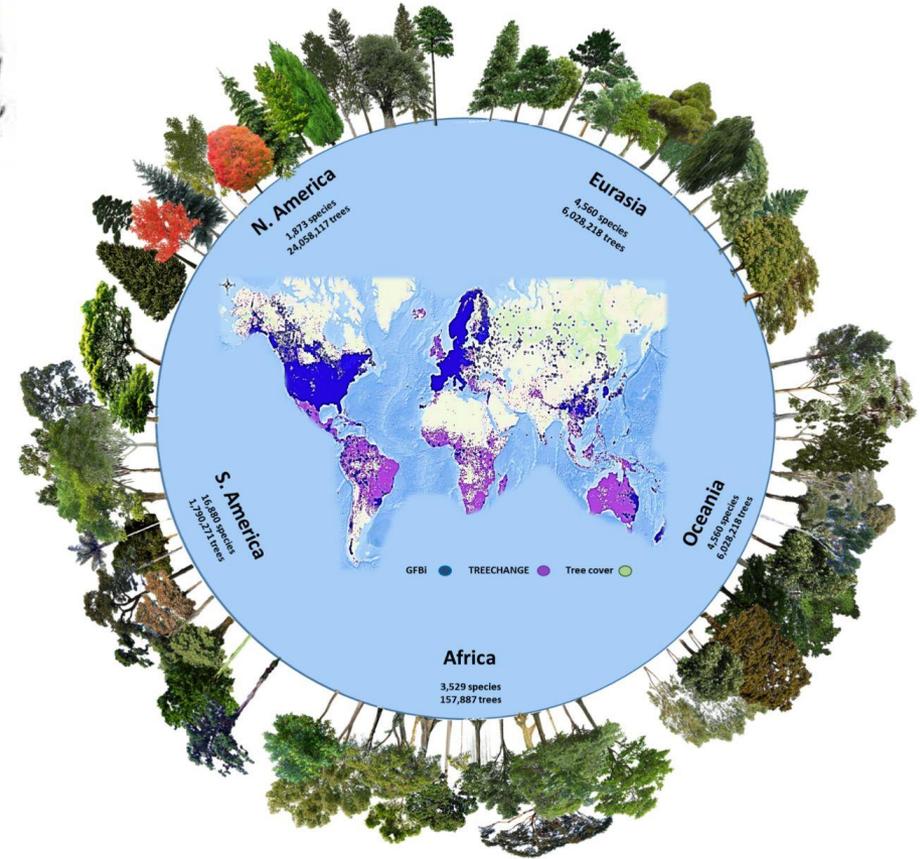
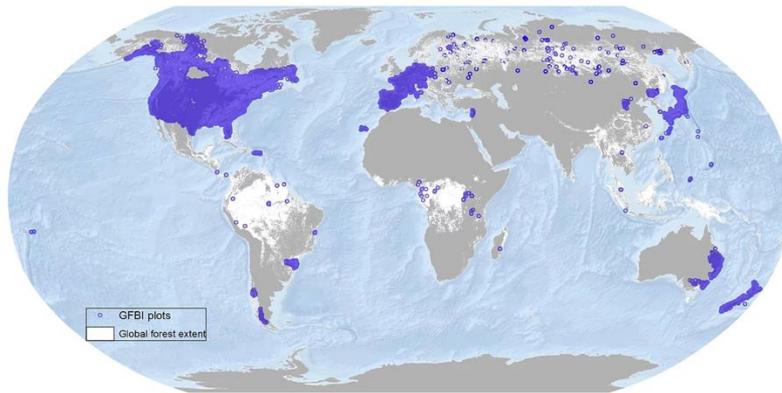
The screenshot shows the PANGAEA Data Publisher homepage. On the left, there is a vertical navigation bar with 'TOPICS' and 'MAP' buttons. The main content area features a 'Submit Data' button with a globe icon, a 'Welcome to PANGAEA® Data Publisher' message, and a search bar. Below the search bar is a grid of 10 topic categories, each with a representative image and a count of data items.

Topic	Count
CHEMISTRY	193458
LITHOSPHERE	133183
BIOLOGICAL CLASSIFICATION	33518
ATMOSPHERE	30143
PALEONTOLOGY	25652
OCEANS	21795
ECOLOGY	18402
LAND SURFACE	7998
BIOSPHERE	4325
GEOPHYSICS	3919

Repositorios de datos

Específicos

Global Forest Inventory Initiative



~73,000 tree species globally, among which ~9,000 tree species are yet to be discovered.

Cazolla et al. 2022

<https://doi.org/10.1073/pnas.2115329119>

Repositorios de datos

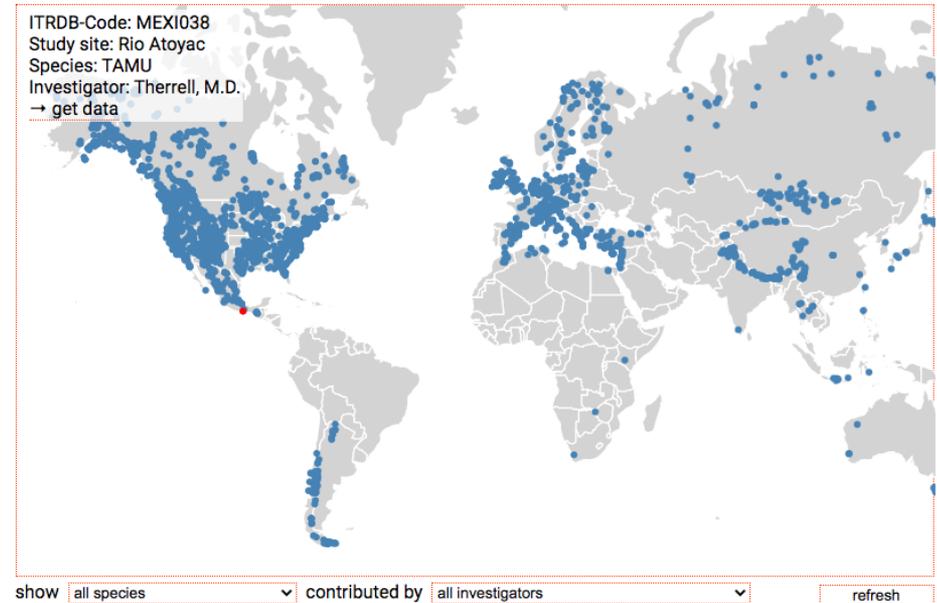
Específicos

International Tree-Ring Data Bank

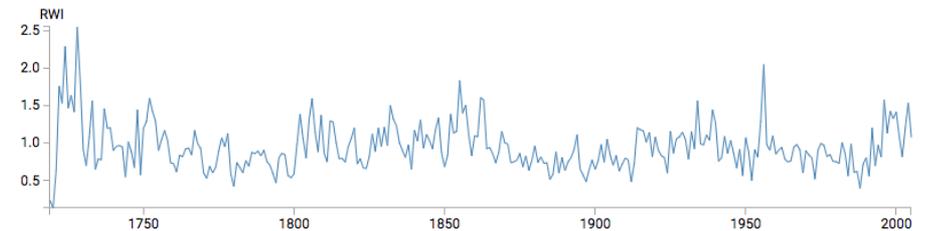
Cronologías de más de 5000 sitios



<http://dendrobox.org/>



CHRONOLOGY



Repositorios de datos

Específicos



Legume Data Portal

22,137,782

Occurrences

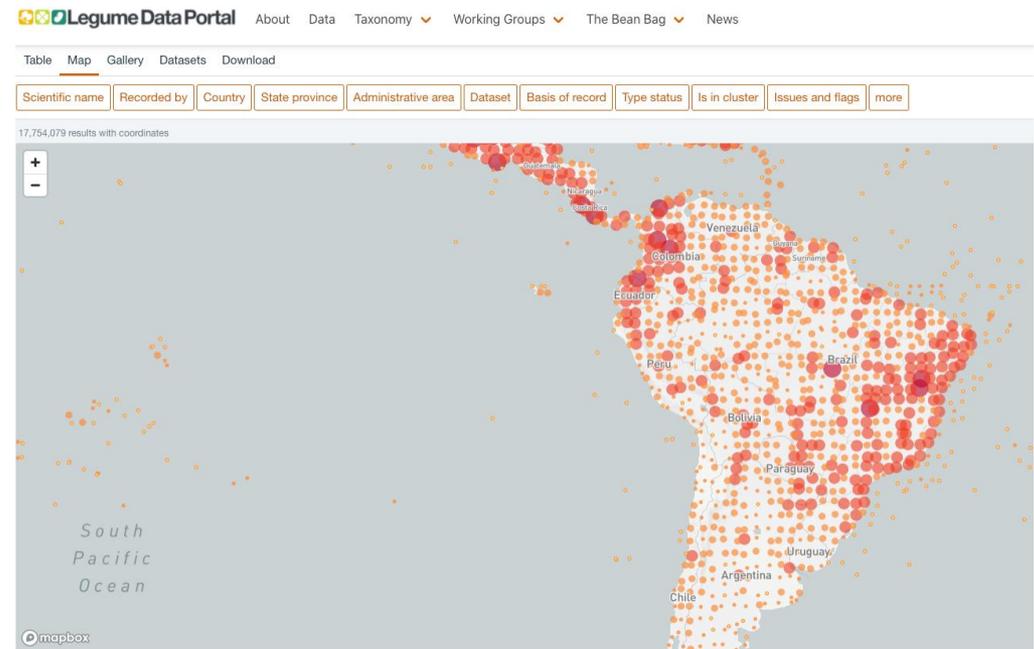
22,360

Species

772

Genera

<https://www.legumedata.org/>

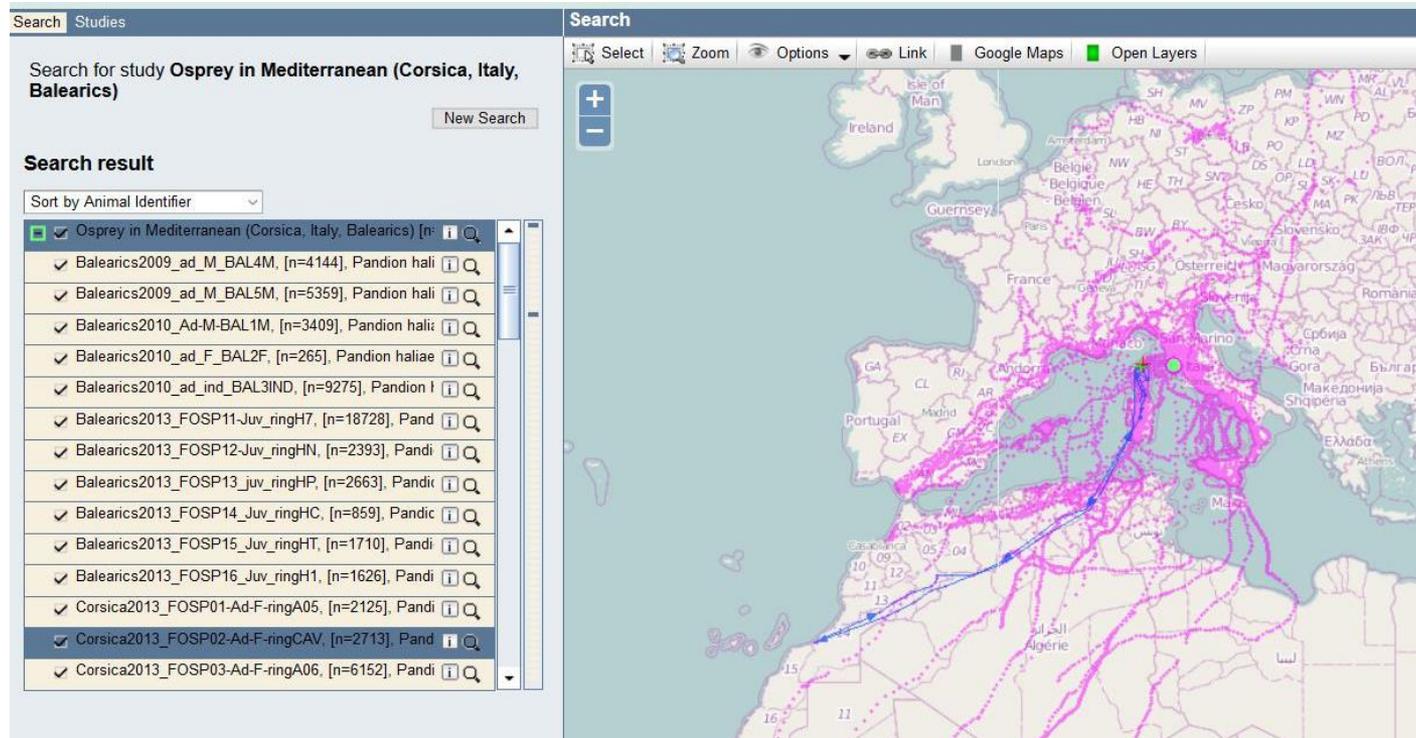


Repositorios de datos

Específicos

Datos de movimiento de animales

<https://www.movebank.org/>



The screenshot displays the Movebank interface. On the left, a search bar contains the text "Search for study Osprey in Mediterranean (Corsica, Italy, Balearics)". Below it, a "Search result" section lists various studies, sorted by "Animal Identifier". The list includes studies from 2009 to 2013, covering different age groups and locations (Balearics, Corsica). The right side of the interface shows a map of Europe with movement tracks overlaid in pink and blue, indicating the migration paths of the tracked animals. The map includes labels for various countries and regions, such as France, Italy, and the Balearic Islands.

Repositorios de datos

Específicos



GBIF

Infraestructura Mundial de Información en Biodiversidad

Es una organización internacional y una red de datos financiada por gobiernos de todo el mundo, destinada a proporcionar a cualquier persona, en cualquier lugar, acceso abierto y gratuito a datos sobre cualquier tipo de forma de vida que hay en la Tierra

La mayor red de datos de biodiversidad del mundo

Nodos y organizaciones participantes + **Secretaría Internacional (Copenhague)**

Objetivo: hacer disponible en internet toda la información que se tiene sobre biodiversidad, para todo el mundo y de manera libre y gratuita.

The GBIF Network

GLOBAL

ÁFRICA

ASIA

EUROPA

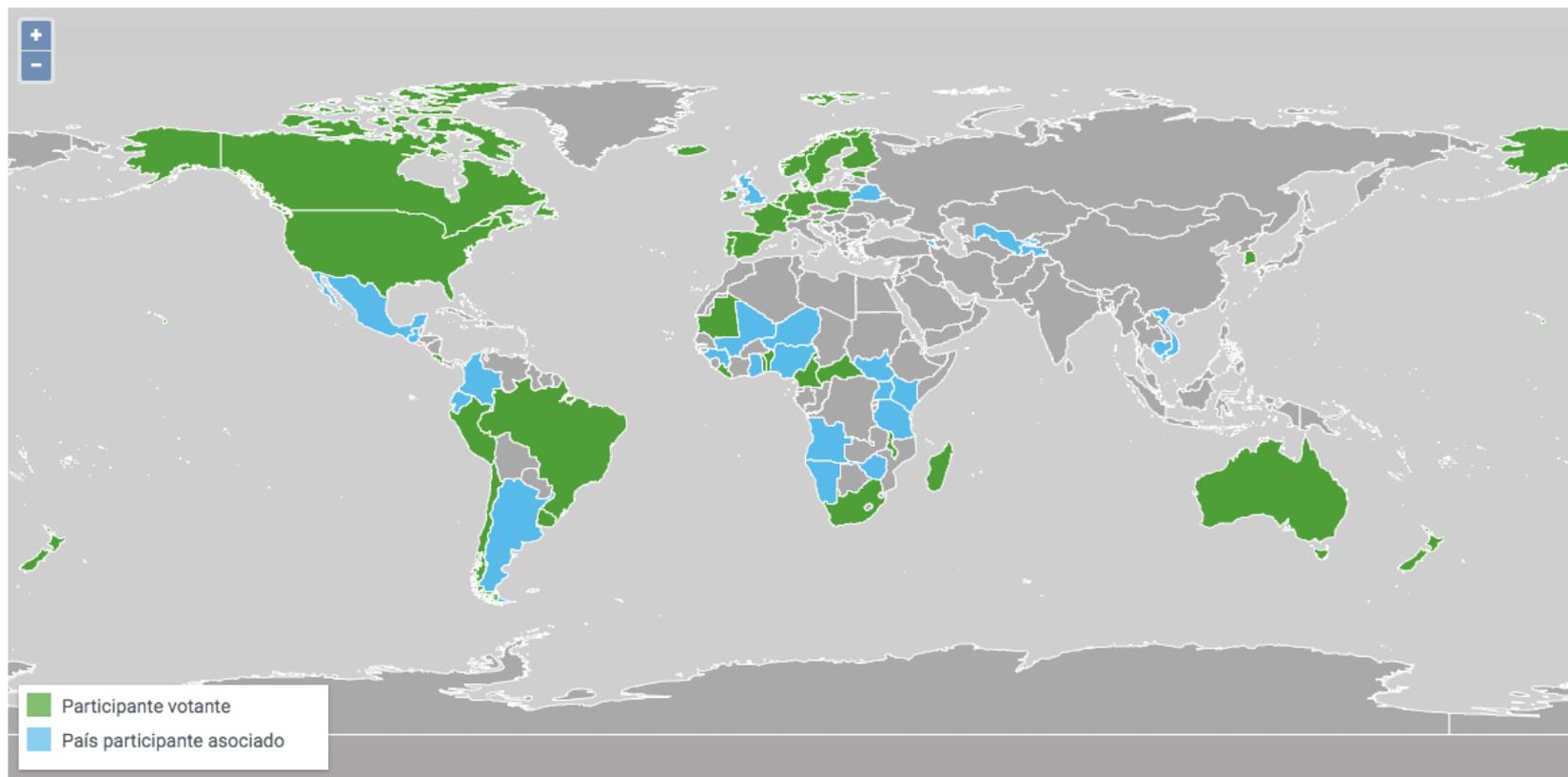
AMÉRICA LATINA Y EL CARIBE

NORTEAMÉRICA

OCEANÍA

ORGANIZACIONES PARTICIPANTES

AFILIADOS A GBIF



39 PARTICIPANTES VOTANTES

24 PAÍSES ASOCIADOS PARTICIPANTES

41 OTROS PARTICIPANTES ASOCIADOS

1,850 PUBLICADORES

Datos How-to Herramientas Community Acerca de

GBIF | Global Biodiversity Information Facility

Acceso libre y gratuito a los datos de biodiversidad

REGISTROS ESPECIES CONJUNTOS DE DATOS PUBLICADORES RECURSOS

[¿QUÉ ES GBIF?](#) [SOBRE GBIF ESPAÑA](#)

Larus michahellis J.F.Naumann, 1840 observed in Monaco by bernardjul

Registros biológicos
1,925,512,011



Noticias

Call for data papers describing datasets from Northern Eurasia

Conjuntos de datos
65,378



Noticias

Call for data papers describing datasets on vectors of human diseases

Instituciones que publican
1,791



Noticias

First thematic help desk to support mobilization of biodiversity data related to disease vectors

Artículos científicos usando datos
6,791



Noticias

New feature enables search of occurrence data by global IUCN Red List Category

Métricas

<https://www.gbif.org/es/analytics/global>

España

Un participante en GBIF con derecho a voto de

Los nombres de los países y áreas están basados en el estándar ISO 3166-1

RESUMEN

DATOS SOBRE

PUBLICACIÓN DE DATOS

PARTICIPACIÓN

ALIEN SPECIES

MÁS...

INFORME DE ACTIVIDAD

DATOS SOBRE

64.456.652

Registros de presencia

2489

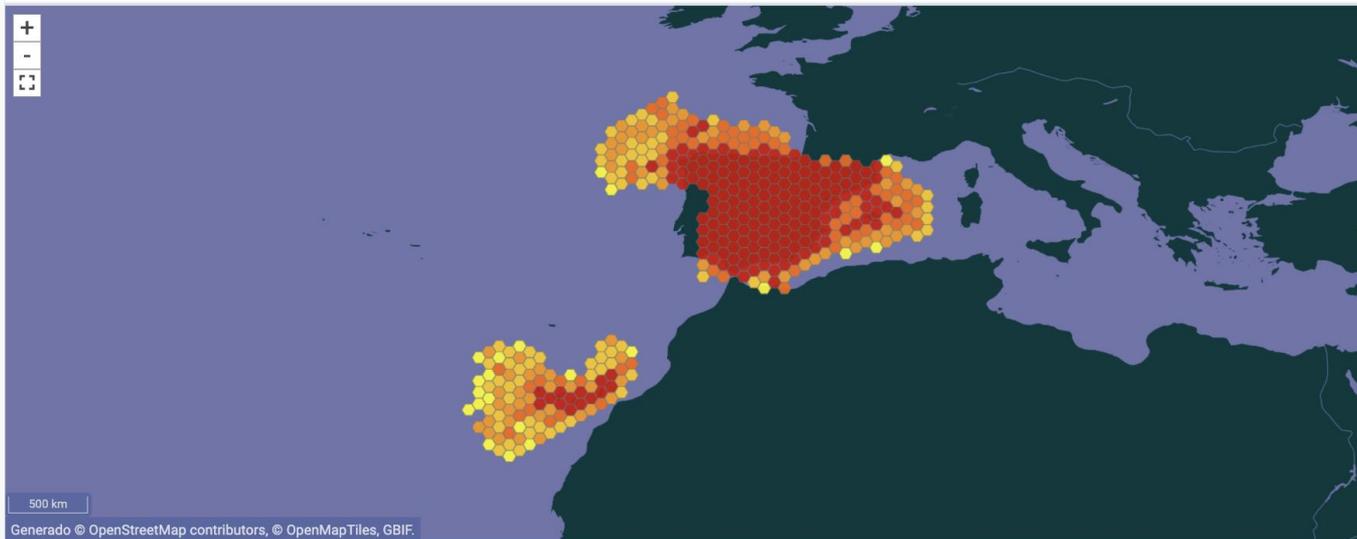
Conjuntos de datos

55

Países y áreas que contribuyen con
datos

637

Publicadores



Cualquier año

1541 - 2024

EXPLORA UN ÁREA



+ info

Comunidad muy activa de aprendizaje

Gran cantidad de recursos

e.g. <https://docs.gbif.org/course-introduction-to-gbif/es/>

Nodo Español <https://elearning.gbif.es/>

Nodo GBIF Colombia <https://biodiversidad.co/>

Acerca de GBIF

<https://docs.gbif.org/course-introduction-to-gbif/es/acerca-de-gbif.html>

¿Alguna duda?

ajpelu.bsky.social

antonio.perez@inia.csic.es

Ayuda JDC2022-050056-I financiada por MCIN/AEI /10.13039/501100011033 y por la Unión Europea NextGenerationEU/PRTR



Si usas esta presentación puedes citarla como:

Pérez-Luque, A.J. (2025). Explorando el ciclo de vida de los datos de biodiversidad y ambientales: desde la recopilación hasta la publicación. Importancia de los datos primarios. Material Docente de la Asignatura: Ciclo de Gestión de los Datos. Master Universitario en Conservación, Gestión y Restauración de la Biodiversidad. Universidad de Granada. <https://ecoinfugr.github.io/ecoinformatica/>