

# Ciencia, datos, resultados e infraestructuras

Francisco Pando  
GBIF.ES (UdC)  
PTI-ECOBIODIV (CSIC)  
Real Jardín Botánico CSIC

Taller online GBIF.ES: Estandarización y publicación de  
datos de biodiversidad de regiones polares. Mayo 2021.



CSIC

Gbif.es



# Preámbulo



Gbif.es

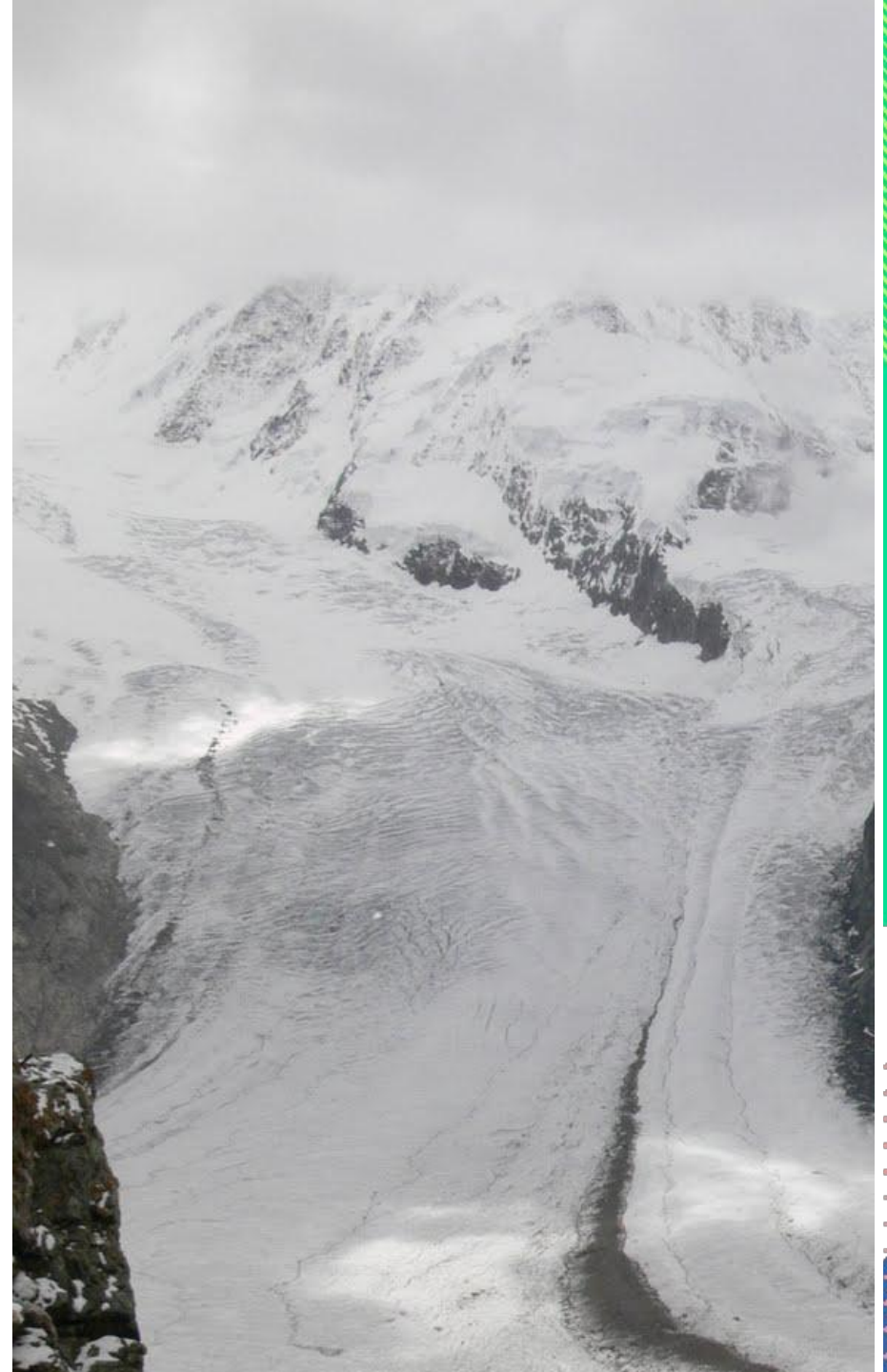


MINISTERIO  
DE CIENCIA  
E INNOVACIÓN



# Contenidos

- El marco del trabajo científico
- Ciencia y datos
- Infraestructuras científicas: antes y ahora
- Publicar datos
- Contexto polar
- Alguna reflexión (polar)





# El marco del trabajo científico

<u>Proyectos</u>	<u>Infraestructuras</u>
Con principio y fin	Indefinidos en el tiempo
Énfasis en resultados	Énfasis en elementos necesarios para "hacer ciencia"
La publicación	Materiales, instalaciones
"Trabajo de científicos"	"Trabajo de técnicos"

**SCIENCE**  
*as a*  
**PROCESS**

Hull, D. L. 1988. Science as a Process. An Evolutionary Account of the Social and Conceptual Development of Science. The University of Chicago Press, Chicago and London, 586 pp.

La Battaglia di dieci uomini nudi. Antonio del Pollaiuolo, 1460-1465



# Énfasis en resultados

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

**Esto es un problema, reconocido desde hace tiempo**



Costes asociados a la publicación en papel

Competición entre científicos

Trabajo extra, sin incentivos

Otras razones (p.ej. Protección de especies)

# Del énfasis en los resultados a los datos

El avance de las tecnologías de la información y comunicación ha deshecho el impedimento de los costes de producción y diseminación.

⇒ Surgimiento de iniciativas para cambiar la cultura científica:

- Para ganar en eficiencia
- Para incrementar el retorno de la inversión pública
- Para ganar en transparencia y por ende en calidad



# Datos y ciencia

## Data-Intensive Scientific Discovery

- Data intensive. Basic science is data intensive in its own right, but data sources that support basic science are often insufficient to support applications. Localized impacts with global extent, such as intrusion of invasive species, are often difficult for centralized projects with small numbers of researchers to ascertain. **New applications-appropriate sources must be identified**, and new ways of observing (including the use of communities as data gatherers) must be developed.

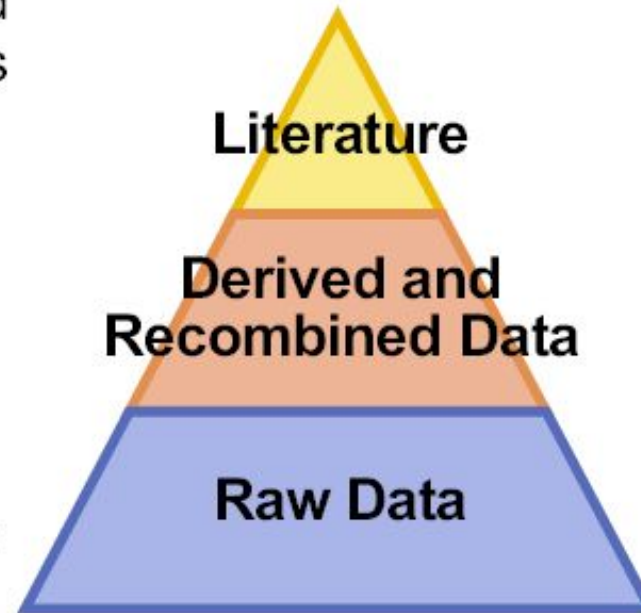
Hey, T., Tansley, S., & Tolle, K. (Eds.) 2009. The fourth Paradigm. Data-Intensive Scientific Discovery. Microsoft Research. Redmon. E.E.U.U.

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>  
<http://serc.carleton.edu/earthandmind/posts/4thpardigm.html>

# De los datos a la ciencia

## All Scientific Data Online

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity





# Datos y política científica

<http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>



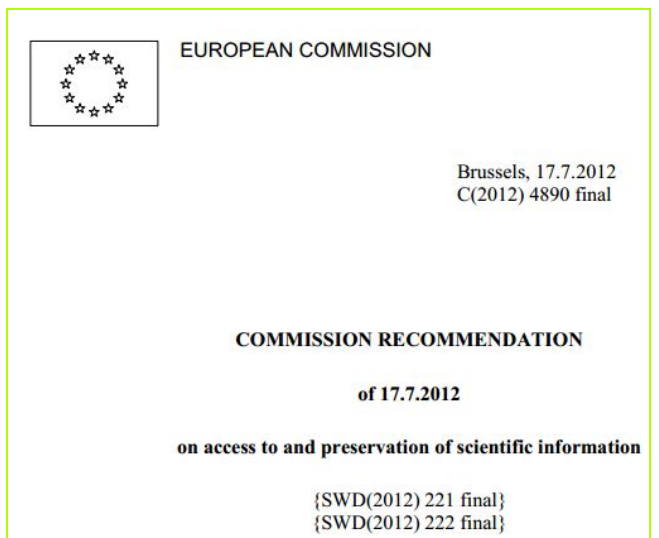
This report recognizes the evolving role of data in science and society and strong and sustainable data sharing and management policies as a critical national need.

## **Digital Research Data Sharing and Management**

December 2011

Task Force on Data Policies  
Committee on Strategy and Budget  
National Science Board

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf)



# Qué se lee ahí

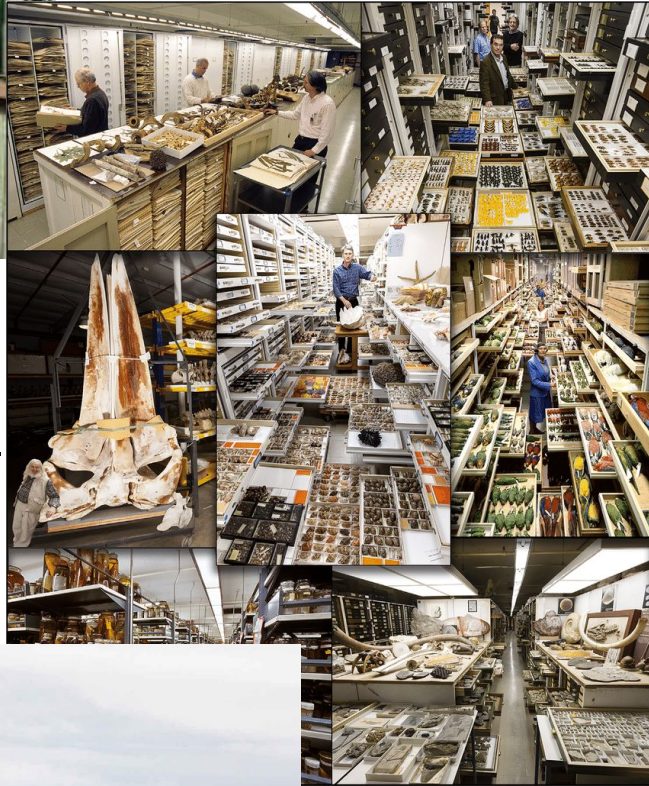
- Sharing can also be encouraged through the establishment of professional incentives such as promoting **the publication of data in a format that allows for citation and verification**.
- Data scientists and curators should be supported by funding agencies and by their home institutions by **providing pathways for advancement to tenure and other reward mechanisms**.
- ...researchers must be confident **when they share data that they will be properly attributed** and the provenance of the data is assured.
- new data **licensing mechanisms** can preserve intellectual property rights and provide researchers with incentives to make their data public.

# Infraestructuras científicas: antes y ahora

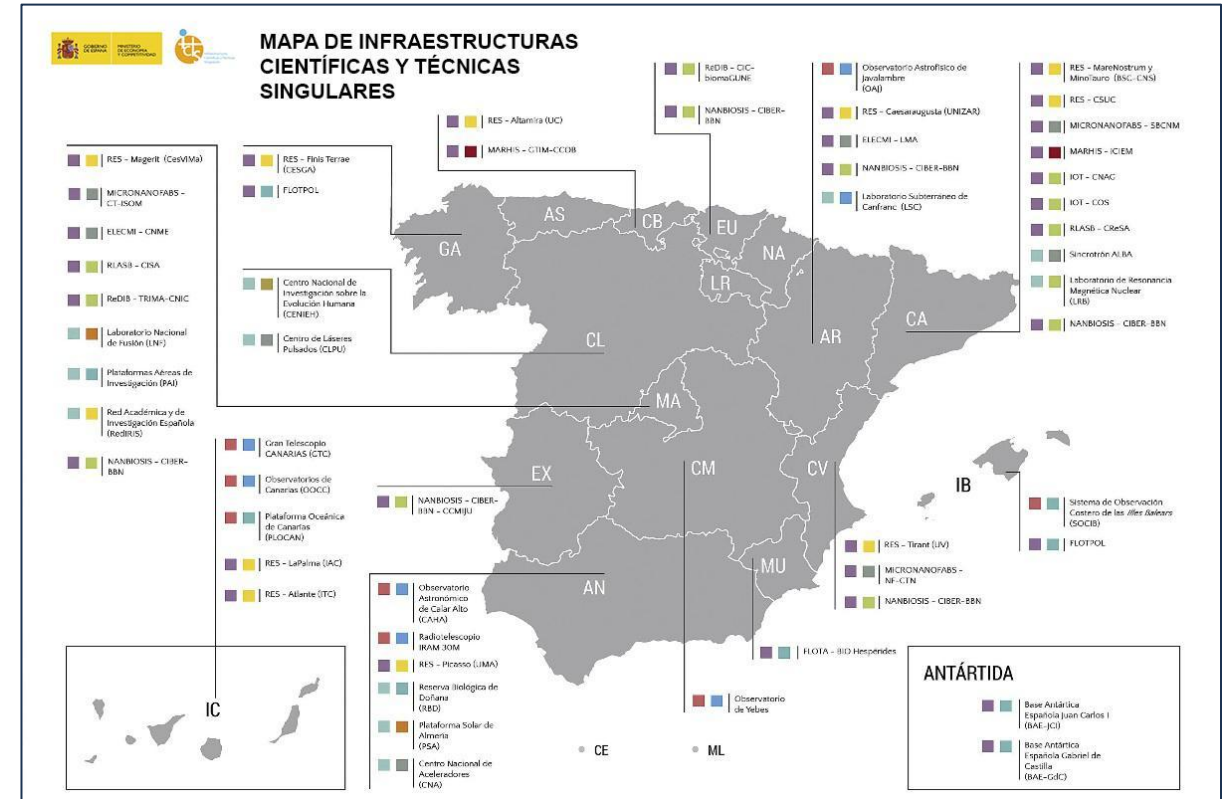


<https://bibdigital.rjb.csic.es/>

<https://naturalhistory.si.edu/>  
Photo Services; assembled  
by A. Tangerini (US)



<http://www.utm.csic.es/>



- Bases Antárticas Españolas (BAEs)
- Flota Oceanográfica Española
- Infraestructura para el cultivo del Atún rojo (ICAR)
- Plataforma Oceánica de Canarias (PLOCAN)
- Reserva Biológica de Doñana (RBD)
- Sistema de Observación Costero de las Illes Balears (SOCIB)

“infraestructuras necesarias para facilitar el acceso de los científicos a entornos naturales que presentan características únicas para la investigación”

<https://www.ciencia.gob.es/portal/site/MICINN/ICTS>



# Infraestructuras científicas: antes y ahora

ESFRI

European Strategy Forum on Research Infrastructures

<https://www.esfri.eu/environment>

DiSSCo (Project)

Distributed System of Scientific Collections

Data-intensive frontier research through unified access to European natural science collections

eLTER (Project)

Long-Term Ecosystem Research in Europe

Integrated cross-scale and cross-disciplinary approach for the analysis of ecosystems and biodiversity

LifeWatch ERIC (Landmark)

e-Infrastructure for Biodiversity and Ecosystem Research

A distributed e-Infrastructure to support research and sustainability of biodiversity and ecosystems

[ACTRIS](#) (Project), [EISCAT\\_3D](#) (Landmark), [EMSO ERIC](#) (Landmark), [ICOS ERIC](#) (Landmark),...

el servicio  
son los  
datos

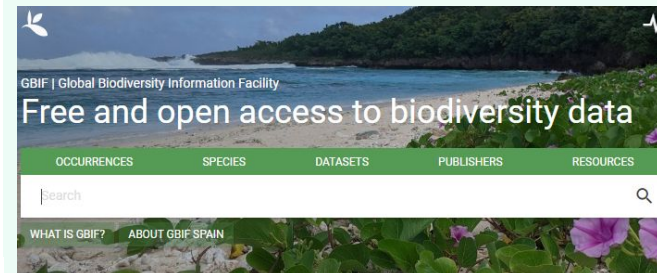
BHL  
Biodiversity Heritage Library

164,468 titles  
267,563 volumes  
59,286,144 pages

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



<http://datos.gbif.es>, <https://www.gbif.es/>, <https://www.gbif.org/>



BY THE NUMBERS | 31 MARCH 2021



# Publicar datos

## Algo de GBIF

- Usa un estándar consolidado de metadatos: EML
- Todos los datos están bajo un estándar común: Darwin Core
- Cada juego de datos mantiene su identidad (sus autores y administradores mantiene el control sobre los mismos)
- Todos los registros están indexados, y permiten que las +57.000 base de datos publicadas se puedan consultar como si fueran una
- Todos los juegos de datos tiene una licencia de datos abiertos “Creative Commons” que son “machine readable”
- Todos los juegos de datos están identificados y son trazables por DOI
- Todas las descargas de datos tiene también su DOI, son públicas y repetibles
- ...

# Datos, ciencia e incentivos

Publicar los datos de manera citable y verificable:  
“artículos de datos”, revistas de datos:



<http://connect.clir.org/blogs/katherine-akers/2013/12/12/data-journals-incentivizing-research-data-dissemination>

- Pensoft data papers:  
<http://biodiversitydatajournal.com/>
- Nature Group: *Scientific Data*.  
<http://www.nature.com/sdata/>

- Datos refrendados por la “revisión por pares”
- Juegos de datos citables
- Juegos de datos identificados por DOIs:
  - accesibles en abierto,
  - trazables,
  - con métricas



# Datos, ciencia e incentivos

scientific data

Explore content | Journal information | Publish with us | Sign up for alerts | RSS feed

nature > scientific data > data descriptors > article

Open Access | Published: 11 April 2017

## A distribution and taxonomic reference dataset of Geranium in the New World

Carlos Aedo & Francisco Pando

Scientific Data 4, Article number: 170049 (2017) | Cite this article

600 Accesses | 1 Citations | 9 Altmetric | Metrics

### Abstract

Geranium L. is a genus of over 350 species distributed throughout most of the world, except in lowland tropical areas. It is the largest genus of the Geraniaceae and is represented in the New World by 137 species. This dataset includes 8,937 records that covers the genus Geranium the New World, providing an updated, taxonomically consistent and a sound geographical distribution of the 137 species of Geranium in America. Specimens from 128 herbaria were reviewed. These were supplemented by others collected during nine field trips, which allowed better knowledge of the variability of characters within populations, and refining species distribution ranges. Each record represents a specimen that has been reviewed and in some cases collected by C. Aedo. Accepted scientific name, locality details, distribution status (introduced, native, naturalized, uncertain), geographic coordinates are given for 8,538 (95%) records, and habitat information for 3,952 (44%). All data have been released under a CC-BY license in a standardized format, which enables easy integration with other data, for example through GBIF.org.

Design Type(s)	data integration objective
Measurement Type(s)	biodiversity assessment objective
Technology Type(s)	digital curation
Factor Type(s)	
Sample Characteristic(s)	Geranium + Americas

Machine-accessible metadata file describing the reported data (ISA-Tab format)

### Background & Summary

In pursuit of preparing a comprehensive world monograph of the genus Geranium, studies of some groups of Geranium from America were carried out as follows:

Species from Canada, Greenland and USA<sup>1,2</sup>.

Sect. *Brasiliensis* R. Knuth from South America<sup>3</sup>.

Figure 1: Map of georeferenced specimens.

Figure 2: Species richness overview.

Figure 3

<https://www.nature.com/articles/sdata201749>

OCCURRENCE DATASET | REGISTERED JULY 12, 2016

## A Distribution and Taxonomic Reference Dataset of Geranium (Geraniaceae) in the New World

Published by CSIC-Real Jardín Botánico

Carlos Aedo • Francisco Pando

8,937 OCCURRENCES | 81 CITATIONS

DATA SET | METRICS | ACTIVITY | DOWNLOAD | HOME PAGE

The dataset includes 8937 records that covers genus Geranium the New World, providing an updated, taxonomically consistent and a sound geographical distribution of the 137 species of Geranium in America. These are the studied specimens on which following monograph was based: Aedo, Carlos (2012). Revision of Geranium (Geraniaceae) in the New World. Systematic Botany Monographs 95: 550p. Besides the exhaustive taxonomic work, a detailed revision of the genus Geranium in America was carried out. The dataset includes 8,937 records that covers the genus Geranium the New World, providing an updated, taxonomically consistent and a sound geographical distribution of the 137 species of Geranium in America. These are the studied specimens on which following monograph was based: Aedo, Carlos (2012). Revision of Geranium (Geraniaceae) in the New World. Systematic Botany Monographs 95: 550p. Besides the exhaustive taxonomic work, a detailed revision of the genus Geranium in America was carried out.

Geranium in America  
CSIC

Publication date: November 30, 2018

Metadata last modified: November 30, 2018

Hosted by: GBIF-Spain

License: CC BY 4.0

How to cite | DOI | 10.15470/ibqzdo

CITATIONS

8,937 Occurrences

95% With coordinates

81 citations

Read more about literature, how it's discovered and linked to GBIF-mediated data.

<https://www.gbif.org/dataset/26d72d3b-4544-4645-aa56-27aa8a669c6f>

ALL LITERATURE

Read more about literature, how it's discovered and linked to GBIF-mediated data.

Biogeography of global drylands

Maestre, F. Benito, B. Berdugo, M. Concomina-Zubiri, L. Delgado-Baquerizo, M. Eldridge, D. ... (2021) New Phytologist

biological soil crusts • diversity • functional traits • macroecology • plant-plant interactions • plant-soil interactions

A global phylogenetic regionalisation of vascular plants reveals a deep split between Gondwanan and Laurasian biotas

Corte, A. Peruzzi, L. Ramirez-Barahona, S. (2021) bioRxiv

Existing global regionalisation schemes for plants consider the compositional affinities among biotas, but these have not considered phylogenetic information explicitly. Incorporating phylogenetic information may substantially advance our understanding of the relationships among regions and the syno...

Working paper

On the Emergence of Biodiversity: Mechanistically Bridging Ecology, Evolution and Paleoenvironments

Hagen, O. (2021) Understanding Darwin and the roles of control

18,445 download events

473 OCCURRENCES FROM THIS DATASET

DOI | 10.15466/dl.lgtrax5

Date: 5 May 2021

Format: Species list

And

- Geometry: POLYGON ((29.5 0, 32 0, 32 -2, 29.5 -2, 29.5 0))
- Has coordinate: true
- Has geospatial issue: false

473 OCCURRENCES FROM THIS DATASET

DOI | 10.15466/dl.a2tqmb

Date: 5 May 2021

Format: Darwin Core Archive

And

- Geometry: POLYGON ((29.5 0, 32 0, 32 -2, 29.5 -2, 29.5 0))
- Has coordinate: true
- Has geospatial issue: false

80 OCCURRENCES FROM THIS DATASET

DOI | 10.15466/dl.mq7zqc

Date: 5 May 2021

Format: Darwin Core Archive

And

- Scientific name: Aves • Tracheophyta • Actinopterygii • Mammalia

# Publicar datos

## Metadatos

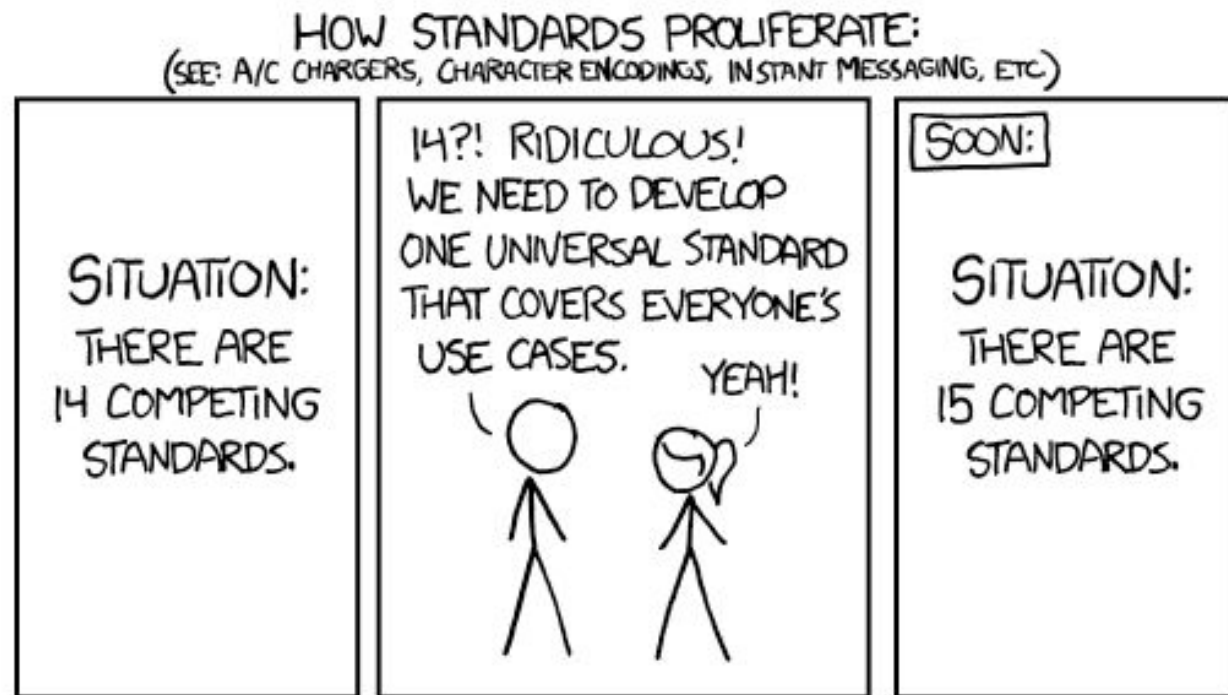
- El conocimiento –contexto si se prefiere– es algo muy difícil de aprehender pero vital para un uso óptimo de los datos
- La documentación proporciona contexto.
- Las tecnologías informáticas están dando acceso a datos e información a una escala inédita hasta ahora, haciendo que esta sea accesible sin importar la distancia o el ámbito donde la información fue generada.
- Este beneficioso fenómeno provoca sin embargo una descontextualización de los datos.
- Esta situación hace que una buena documentación [metadatos] sea más importante que nunca.

HOW I  
LEARNED  
TO STOP  
WORRYING  
AND  
LOVE  
METADATA

# Publicar datos

## Estándares

- “La misma información es registrada del mismo modo”
- Varios tipos, los interesantes aquí, “especificaciones de datos/perfil de datos” > Darwin Core > Dublin Core
- No inventar la rueda
- Es mejor un estándar ampliamente usado que uno más rico de menor implantación
- Un estándar de intercambio no debe ser tomado como una norma de implementación en un sistema de gestión/producción





# Lo polar en GBIF

## Scientific Committee on Antarctic Research

Associate participant

Member since: 2008-02 • Country or area: United Kingdom of Great Britain and Northern Ireland • [Website](#)

105 datasets

### SCAR - AntOBIS

Joined 4 years ago

AntOBIS is the Antarctic Thematic Node of OBIS. SCAR, the Scientific Committee on Antarctic Research, is an interdisciplinary body of ICSU, the International Council for Science, and it is charged wit...



87 datasets 128 citations United Kingdom of Great Britain and Northern Ireland

### SCAR - Expert Group on Birds and Marine Mammals

Joined 8 years ago

The new SCAR EG-BAMM is an exciting initiative tasked with providing expert knowledge and research leadership in all matters related to birds and mammals in the Antarctic, in order to support research...

Antarctica

### Scientific Committee on Antarctic Research - Marine Biodiversity Information Network (SCAR-MarBIN)

Joined 11 years ago

[www.scarmarbin.be](http://www.scarmarbin.be)

### Wichita State University Invertebrate Collection

Joined 2 years ago

The Wichita State University Biodiversity Laboratory houses the invertebrate collection (WICHI) which contains arthropod specimens that result from regional research (pollination, herbivory, habitat s...

1 dataset 5 citations United States of America

### SCAR - Microbial Antarctic Resource System

Joined 7 years ago

83 datasets 15 citations

<https://www.gbif.org/participant/306>

Biological Institute Helgoland (BAH) in the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research

8,603 OCCURRENCES 8,603 HOSTED OCCURRENCES 1 DATASET 65 CITATIONS

<https://www.gbif.org/publisher/497688a0-59d6-11db-893e-b8a03c50a862>

Norwegian Polar Institute

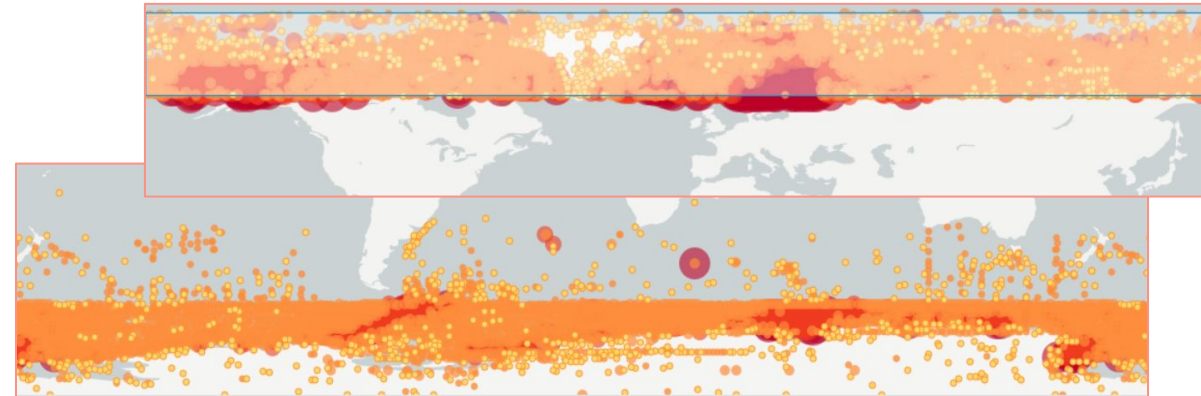
<https://www.gbif.org/es/publisher/ad07600d-5425-43e2-9175-0353bb6d40a2>

UK Polar Data Centre

<https://www.gbif.org/publisher/80420c96-95d0-44eb-9f77-339ac92051fb>

Polar-Alpine Botanical Garden-Institute of N.A. Avrorin KSC RAS

<https://www.gbif.org/publisher/2d52b584-21e1-4265-9c56-8cb5daf45181>





**International Forum on 'Polar Data Activities in Global Data Systems'**  
National Museum of Nature and Science, Tokyo  
**15–16 October 2013**

## HARMONIZING POLAR BIODIVERSITY DATA FOR WIDER ACCESS AND INTEGRATION: A COLLABORATION BETWEEN THE SPANISH POLAR DATA CENTER AND GBIF-SPAIN

O. Bermúdez<sup>1\*</sup>, V. González-Alvaro<sup>2</sup>, F. Pando<sup>2</sup>, A. Barragán<sup>1</sup>, K. Cezón<sup>2</sup>, C. Lujano<sup>2</sup>, S. Martínez de la Riva<sup>2</sup>, C. Villaverde<sup>2</sup>, P. Ríos<sup>3</sup>  
<sup>1\*</sup> National Polar Data Centre (NPDC), Spanish Geological Survey, Ríos Rosas 23, Madrid, 28003, Spain  
<sup>2</sup> GBIF Spain, Coordination Unit, Royal Botanical Garden-CSIC, Plaza Murillo 2, Madrid, 28014, Spain  
<sup>3</sup> Oceanographic Center of Gijón, Spanish Institute of Oceanography (IEO), Avda Príncipe de Asturias 70bis, Gijón, 33212, Spain  
 Email: o.bermudez@igme.es

The Global Biodiversity Information Facility (GBIF) is an intergovernmental organization that promotes and facilitates the mobilization, access, discovery and use of information about organisms over time and across the planet, accomplishing the original purpose of its founding by governments in 2001, i.e. to encourage free and open access to primary biodiversity data over the Internet.

SCAR (Scientific Committee on Antarctic Research) is an Associated Participant of GBIF from 2008. Spain is a full member, and the associated Spanish Polar Data Centre (NPDC) is responsible of the management and upload of data (and metadata) from Spanish projects developed in both the Arctic and Antarctic regions in a, at the same time, national and international database (Antarctic Master Directory, Global Change Master Directory). As a result, the CNPD allows an excellent information flow to boost the polar research.

Here we present the collaborative work between GBIF Spain and NPDC. In the one hand, making polar biodiversity data available through GBIF seemed a logical and effective way to establish proved methods to manage biodiversity polar data. On the other hand, polar biodiversity data published by Spanish Institutions through GBIF can be identified and incorporated into the NPDC.

For those purposes, three datasets were selected in order to establish a methodology that could be later applied to other datasets.

## CONCLUSIONS

A procedure for connecting two large data avenues such as SCAR and GBIF has been defined in a way that can be easily expanded and replicated.

The work carried out also demonstrate to projects dealing with polar biodiversity data how the impact of their research can be multiplied.

**GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)**  
free and open access to biodiversity data

email  password  [Iniciar Sesión](#) [ESPAÑOL](#)

[Inicio](#) [Acerca de](#)

Resumen

Descargas

Versiones

¿Cómo referenciar?

Derechos

Registro GBIF

Palabras Clave

Contactos

[EML](#) [RTF](#) [Versiones](#) [Derechos](#) [Citar](#)

**Abundances of Virus-like particles (VLP) resulted in a mesocosm experiment performed in Lake Limnopolar**

Última versión Publicado por GBIF-Spain en Mar 8, 2021

DOI: [10.15470/abiywl](https://doi.org/10.15470/abiywl)

This data are part of the activities for the accurate description of the limnetic systems in Byers Peninsula (Livingston Island), considered one of the most important freshwater zones in the Maritime Antarctica. Data are the abundances of Virus-like particles at different experimental conditions of a mesocosm assay performed in the Lake Limnopolar with the aim to evaluate the role of both biotic and abiotic interactions in the pelagic community

**GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)**  
free and open access to biodiversity data

email  password  [Iniciar Sesión](#) [ESPAÑOL](#)

[Inicio](#) [Acerca de](#)

Resumen

Registros

Descargas

Versiones

¿Cómo referenciar?

Derechos

Registro GBIF

Palabras Clave

**Ecophysiological and floristic study of the lichen communities in Livingston Island**

Última versión Publicado por GBIF-Spain en Mar 8, 2021

DOI: [10.15470/idn6aw](https://doi.org/10.15470/idn6aw)

Scientific objectives: -Floristic study of Livingston Island. -Structure, succession and dynamism of the principal lichen and bryolichen communities. -Chronological study (distribution) in the Juan Carlos I base area of plantspecies with special bioclimatic significance. -Research on the early development of lichen symbiosis in relation to microclimate and different rock types. -Colonization and adaptation strategies of lichens in the Antarctic environment. Activities carried out: The Byers Peninsula was explored. In this extensive ice-free region, a largenumber of samples were collected on acid to ultrabasic rocks, both eruptive and sedimentary. The following places were visited: 1) Beach and front moraine

## Alguna reflexión (polar... o no)

Interacciones entre participantes (vosotros)

Invitación a “extender la palabra”

El juego de los datos en ciencia es de dos direcciones, publicar y usar, y en las dos importa ser honesto y mantener el pensamiento crítico



Francisco Pando

Responsable

GBIF.ES, Unidad de Coordinación

Real Jardín Botánico - CSIC

Plaza de Murillo, 2

28014Madrid, Spain

[pando@gbif.es](mailto:pando@gbif.es)



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

**PTI-ECOBIODIV**, Plataforma Temática Interdisciplinar del CSIC para la **Síntesis de Datos de Ecosistemas y Biodiversidad**, tiene como misión la construcción de una plataforma de ciencia abierta de publicación e integración de datos de biodiversidad y ambientales, con capacidades básicas de análisis y de desarrollar laboratorios virtuales (con capacidades avanzadas de análisis, síntesis y predicción) sobre la misma. <https://pti-ecobiodiv.csic.es/>

**PTI-POLARCSIC**: Plataforma Temática Interdisciplinar: “**Observatorio de zonas polares: Horizonte 2050**” viene justificada, tanto por el claro interés nacional/internacional en la investigación polar como por los numerosos grupos del CSIC de diversas disciplinas que trabajan en la Antártida y Ártico de forma activa y con resultados científicos de relevancia. <https://polarcsic.es/>

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

