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.

MINISTERIO DE CIENCIA E INNOVACIÓN
CSIC
Gbif.es
PTI ECOBIODIV CSIC
Consejo Superior de Investigaciones Científicas
POLAR CSIC

Preámbulo

GBIF
Global Biodiversity Information Facility

Gbif.es

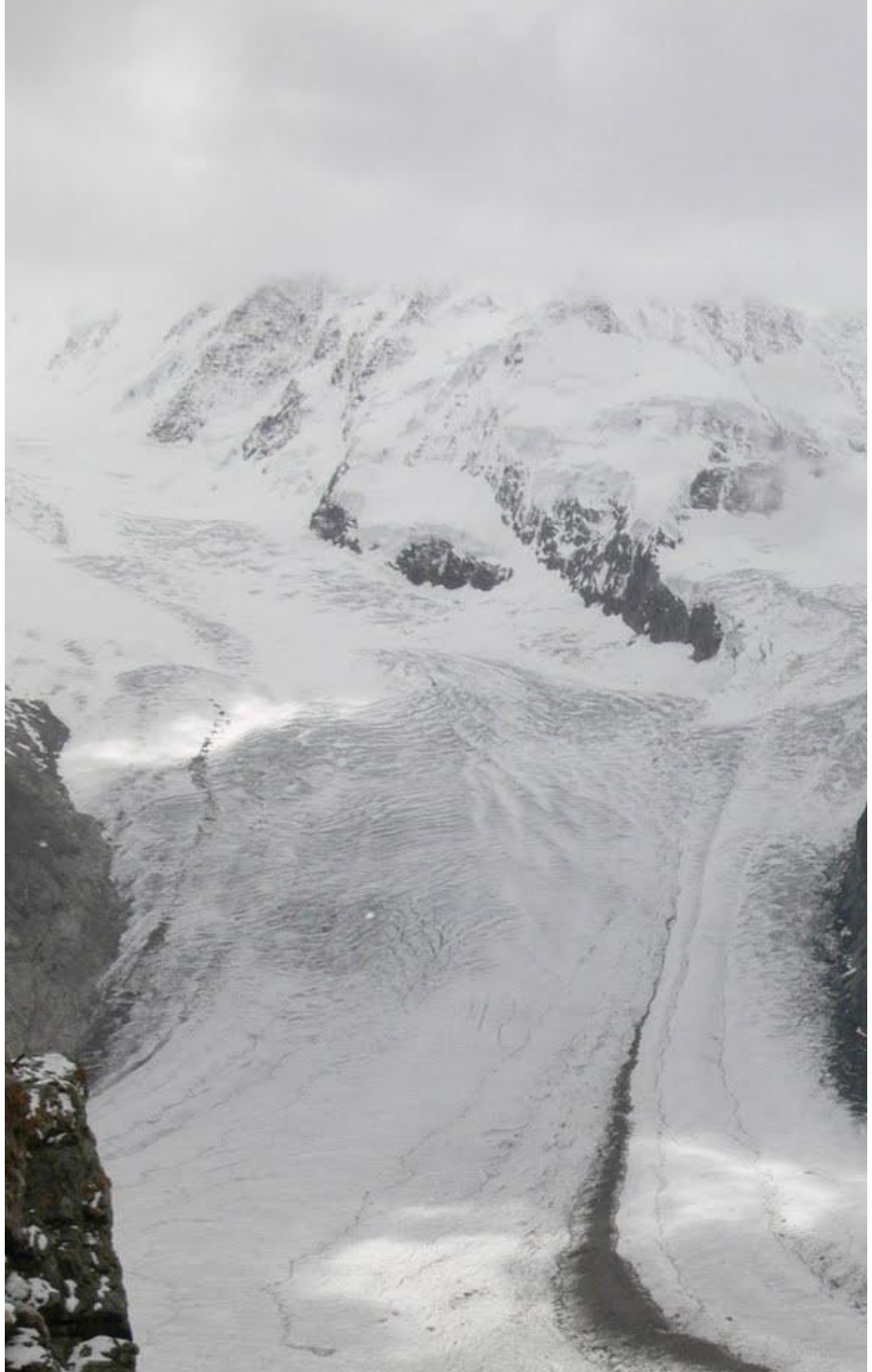
POLAR
CSIC

MINISTERIO DE CIENCIA E INNOVACIÓN

PTI ECOBIODIV
CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

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

Proyectos	Infraestructuras
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"

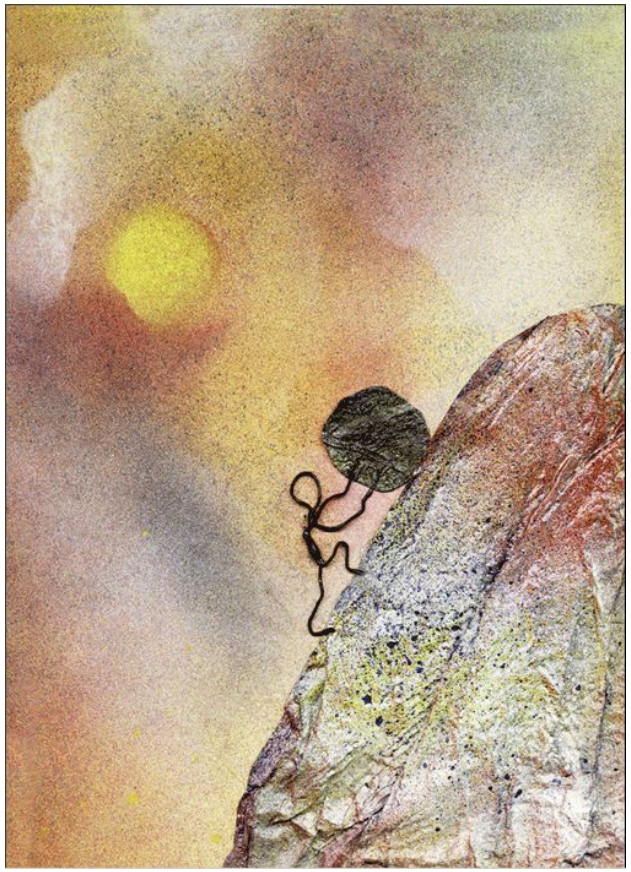
Hull, D. L. 1988. Science as a Process. An Evolutionary Account of the Social and Conceptual Development of Science. The University of Chicago Press, Chicago and London, 586 pp.

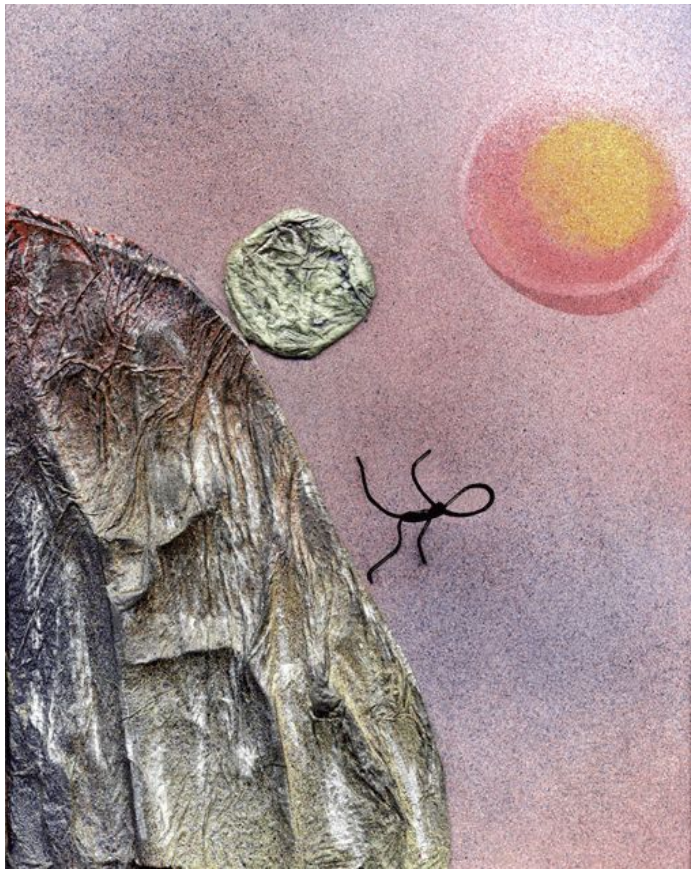
La Battaglia di dieci uomini nudi. Antonio del Pollaiolo, 1460-1465

Énfasis en resultados

HEYWOOD, V.H. (1974). Systematics-the stone of Sysyphus. 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.

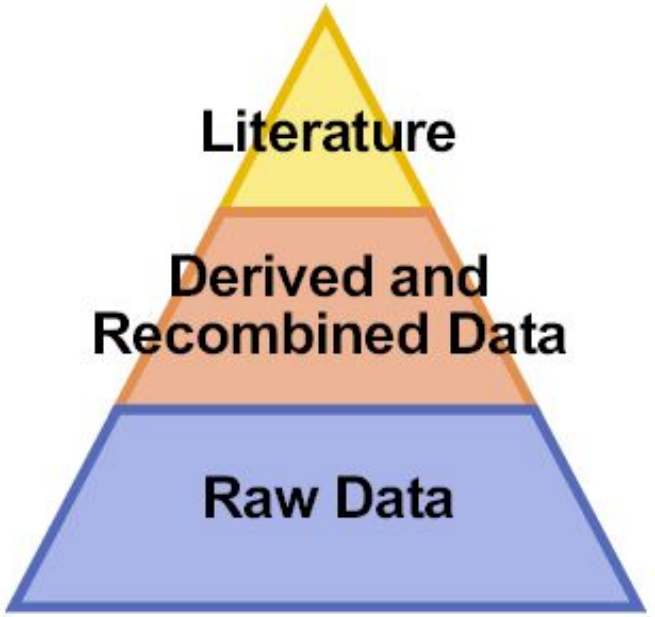
Hey,T., Tansley,S., & Tolle,K. (Eds.) 2009. The fourth Paradigm. Data-Intensive Scientific Discovery. Microsoft Research. Redmon. E.E.U.U.

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

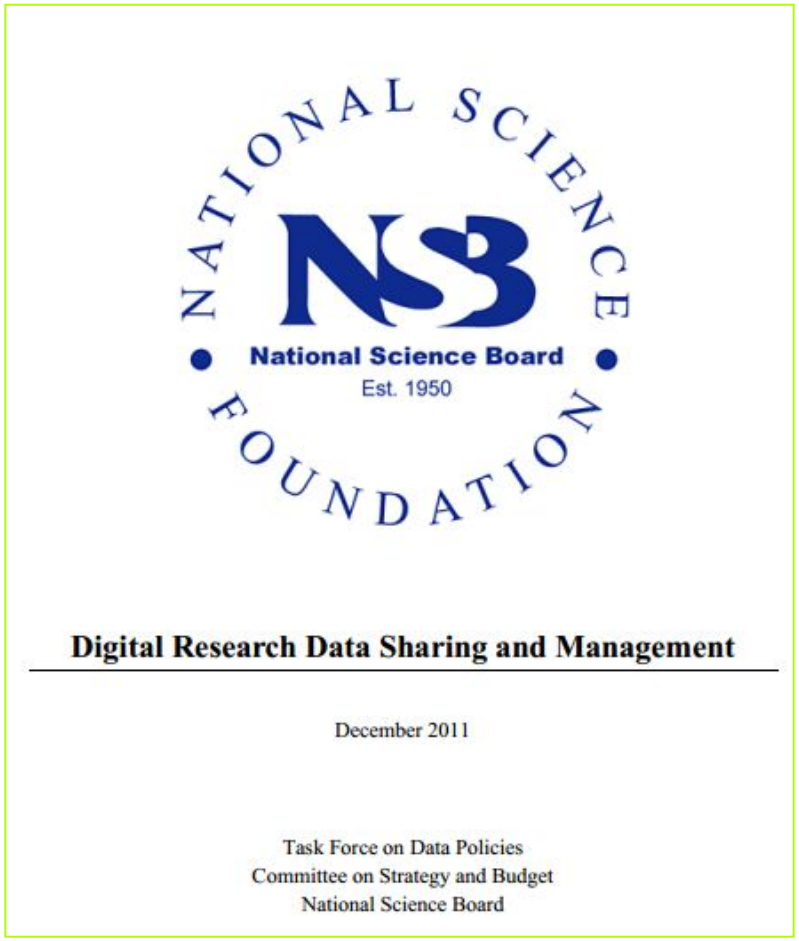
- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity



The Fourth Paradigm: Data-Intensive Scientific Discovery
http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx

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.

http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

Qué se lee ahí

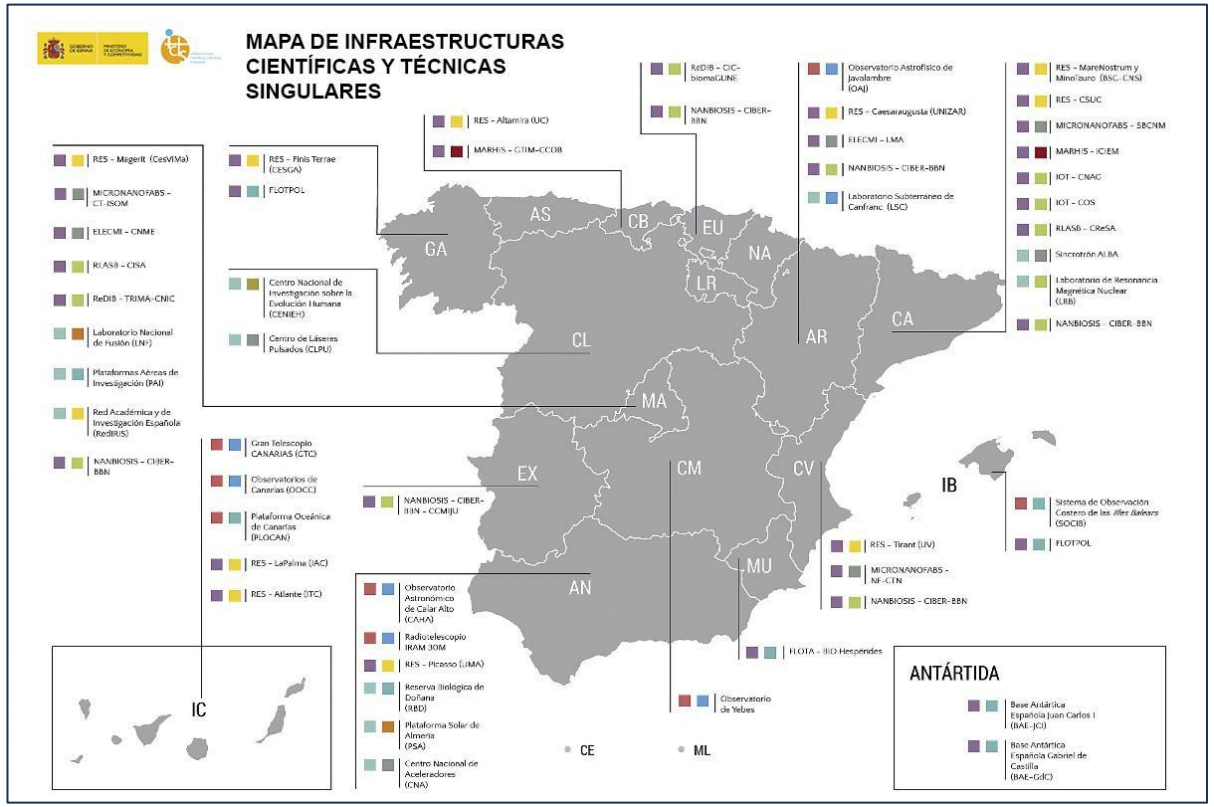
- Sharing can also be encouraged through the establishment of professional incentives such as promoting the publication of data in a format that allows for citation and verification.
- Data scientists and curators should be supported by funding agencies and by their home institutions by providing pathways 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),..

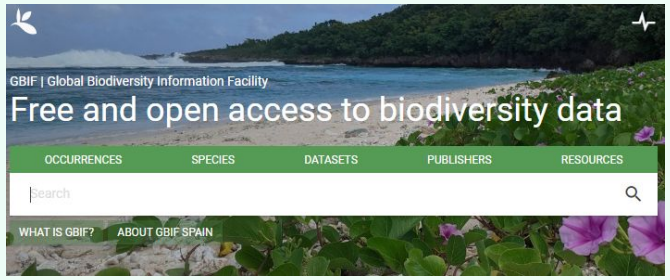
el servicio son los datos

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/



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

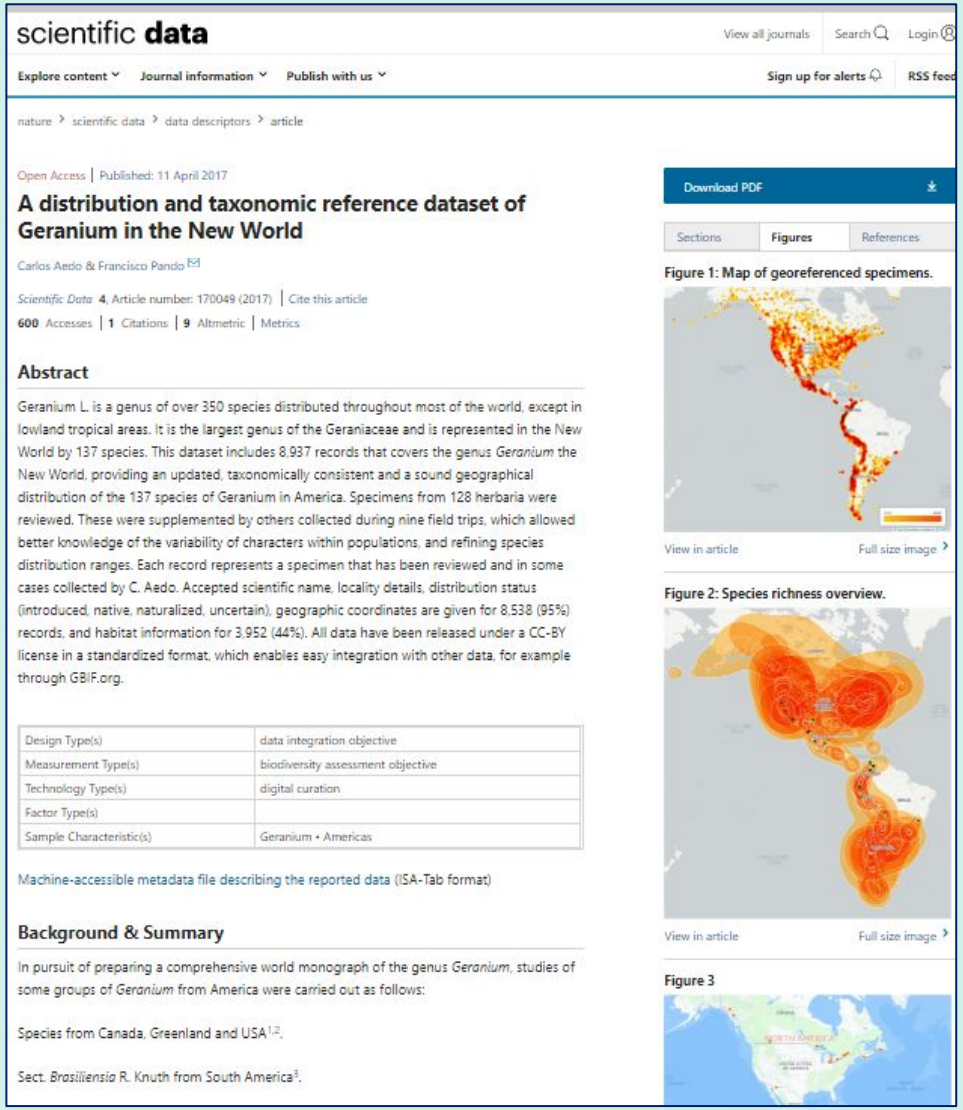
Publicar los datos de manera citable y verificable:
"artículos de datos", revistas de datos:

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

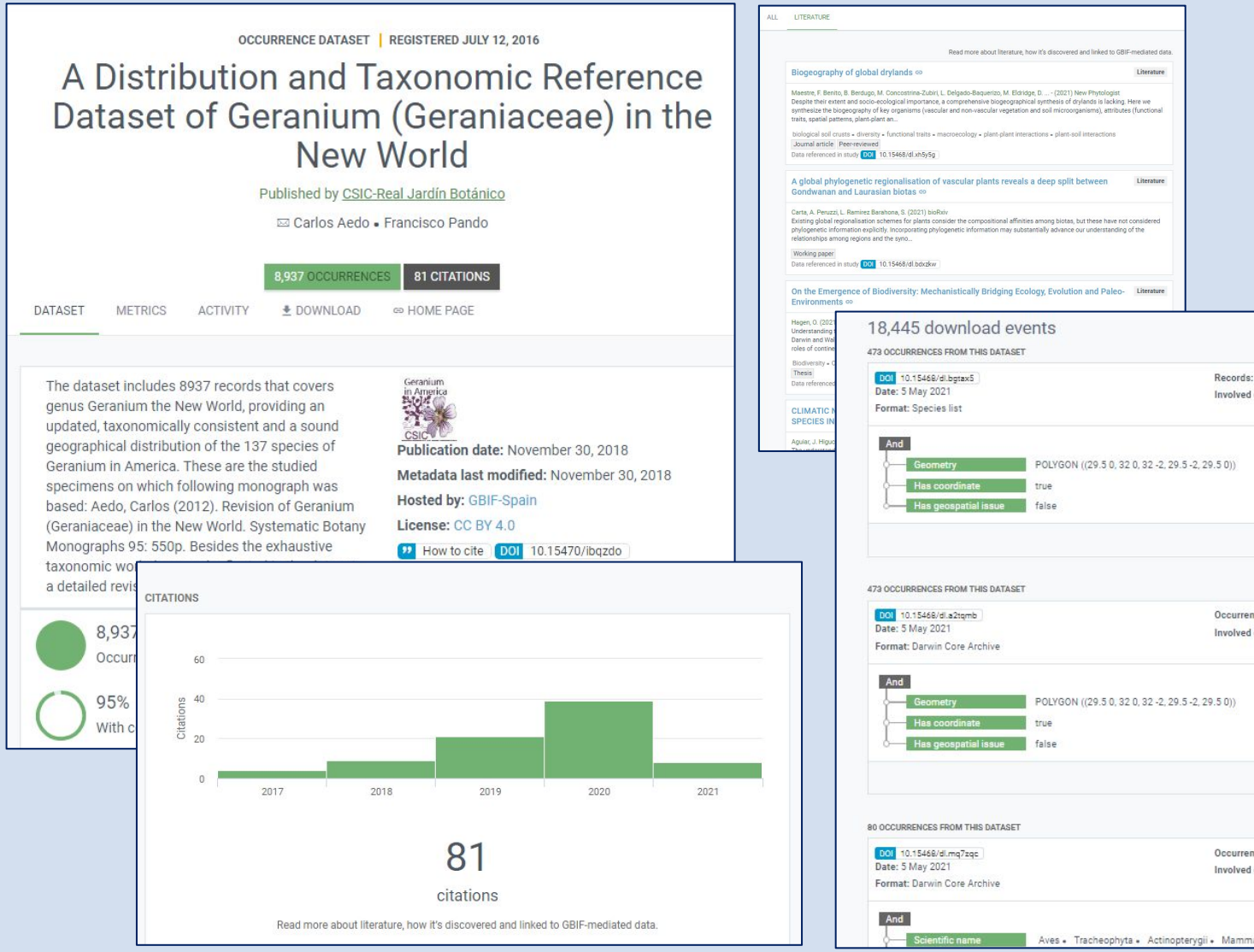
- Pensoft data papers:
  http://biodiversitydatajournal.com/
- Nature Group: *Scientific Data.*
  http://www.nature.com/sdata/

- 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



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



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

Publicar datos

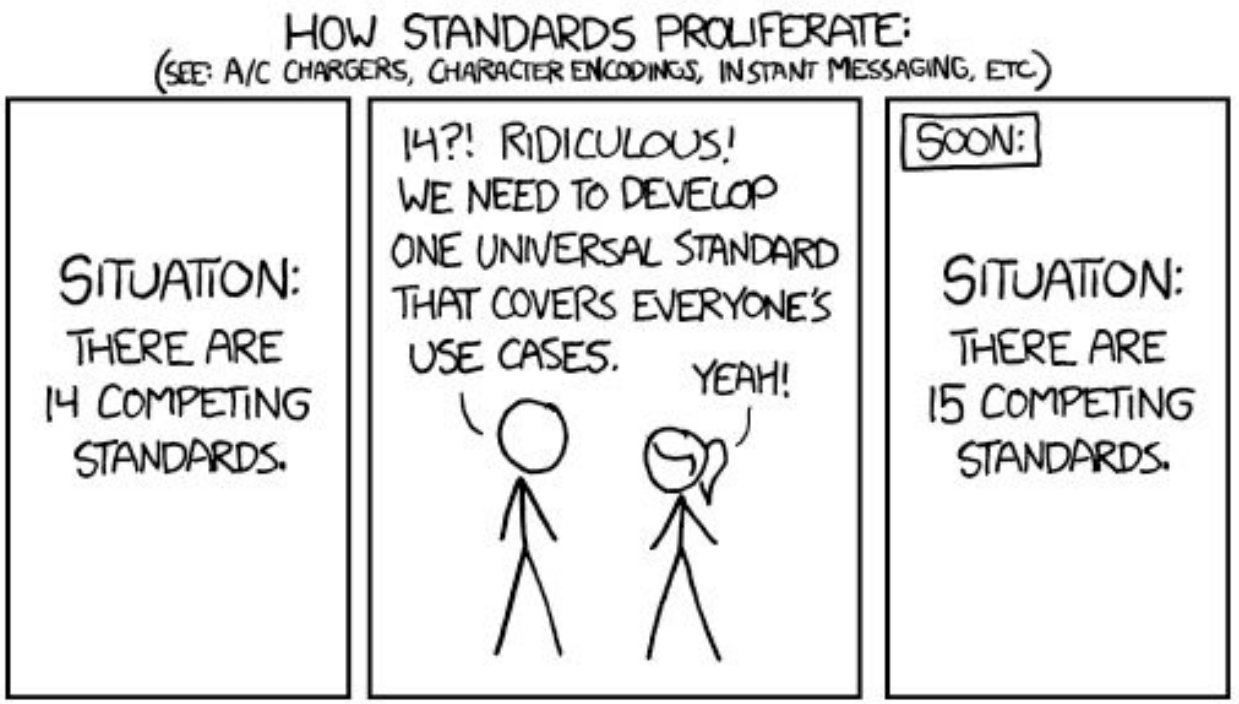
Metadatos

- 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.

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



https://xkcd.com/927/ CC-BY-NC 2.5

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

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

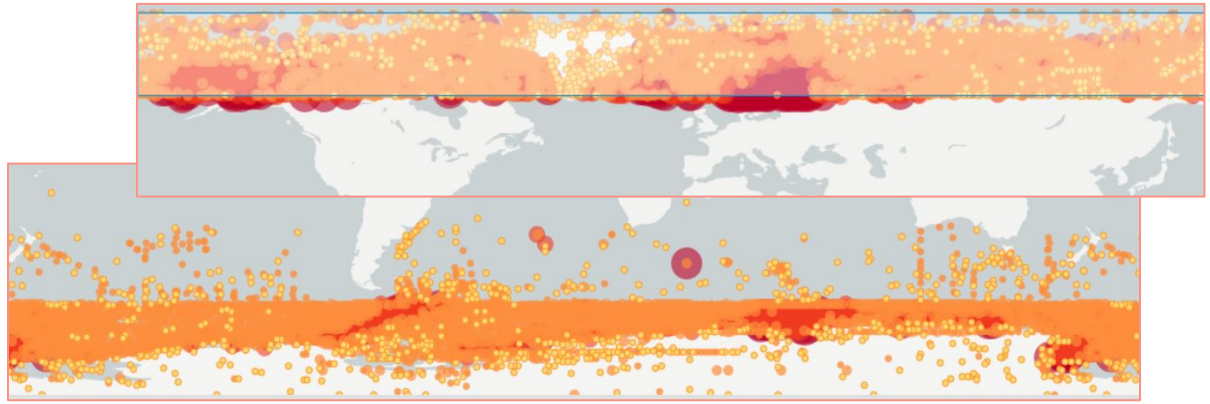
8,603 OCCURRENCES | 8,603 HOSTED OCCURRENCES | 1 DATASET | 65 CITATIONS

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. Bermudez[1], V. Gonzalez-Alvaro[2], F. Pando[2], A. Barragán[1], K. Cezon[2], C. Lujano[2], S. Martinez de la Riva[2], C. Villaverde[2], P. Rios[3]*
[1] National Polar Data Centre (NPDC), Spanish Geological Survey, Rios Rosas 23, Madrid, 28003, Spain*
[2] GBIF Spain, Coordination Unit, Royal Botanical Garden-CSIC, Plaza Murillo 2, Madrid, 28014, Spain
[3] Oceanographic Center of Gijon, Spanish Institute of Oceanography (IEO), Avda Principe de Asturias 70bis, Gijon, 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 (NPDT) 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 CNDP allows an excellent information flow to boost the polar research.

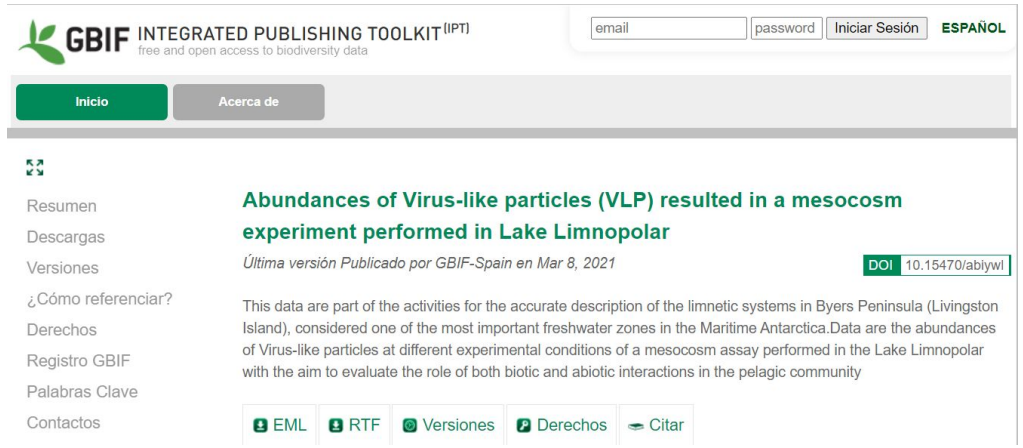
Here we present the collaborative work between GBIF Spain and NPDT. In the one hand, making polar biodiversity data available through GBIF seemed a logical and effective way to establish proved methods to manage biodiversity polar data. On the other hand, polar biodiversity data published by Spanish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

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.



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

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/

MINISTERIO DE CIENCIA E INNOVACIÓN

CSIC

Gbif.es

POLAR CSIC

PTI ECOBIODIV CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS