

Datos buenos y datos malos (de biodiversidad)

Francisco Pando

Unidad de Coordinación, nodo español
de GBIF (Global Biodiversity
Information Facility) / CSIC



EUSKO JAURLARITZA
GOBIERNO VASCO

Foro Social de la Biodiversidad de Euskadi 2024-2025

Dpto. Departamento de Industria, Transición Energética y Sostenibilidad
Medio Ambiente

Donostia, 24 de enero de 2025

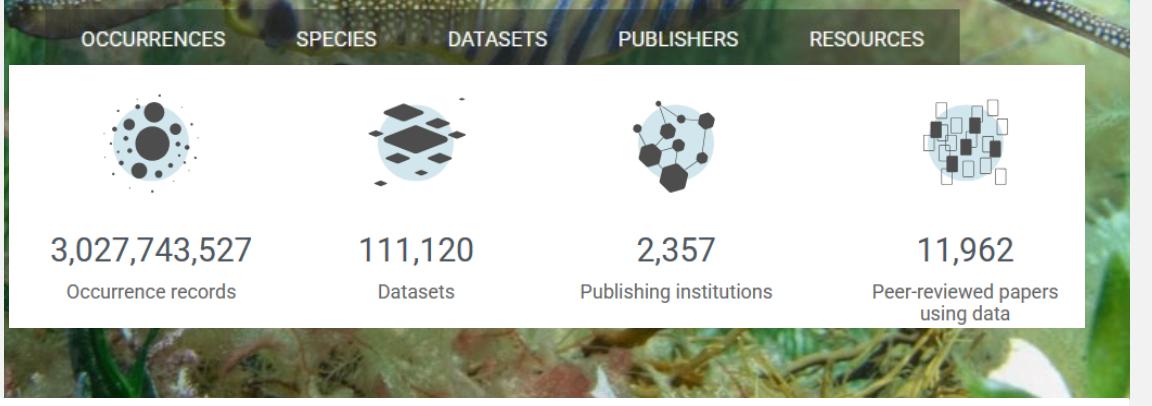




Qué vamos a ver

- Por qué esta charla
- GBIF en una diapositiva
- La ciencia y los datos
- El imperio de los datos
- Qué hace un dato bueno
- + qué hace un dato bueno
- Mención especial: identificadores
- Dos colaterales
- Principios FAIR
- La última diapo

Free and open access to biodiversity data



PUBLISHER | SINCE JULY 27, 2015

Basque Government

113,232 OCCURRENCES | 78,151 HOSTED OCCURRENCES | 66 DATASETS | 307 CITATIONS

ABOUT | METRICS | [HOME PAGE](#)

Description: The Basque Government is the governing body of the Basque Autonomous Community of Spain. Its main responsibilities on nature and biodiversity are: • Creation of legally binding protected natural areas, • Definition, promotion and administration of Nature 2000 areas within the Basque Autonomous Community, • Planning of the high ecological valued and fragile ecosystems, • Administration and management of the Basque Catalogue of Threatened Species, • Elaboration of plans for managing marine and land natural resources, • Elaboration of studies, plans and strategies for the conservation and protection of awareness in matters of environment • Management of envi

Endorsed by: [GBIF Spain](#)

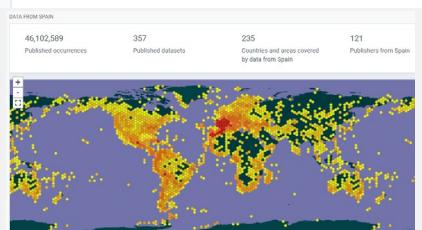
Installations: [Basque Government installation](#)

Administrative contact: [Marta Iturribarria](#)

Technical contact: [Marta Iturribarria](#)

Country or area: [Spain](#)

Hosting: [61 datasets](#) (1 publisher • 1 country)



Nodo Nacional de Información en Biodiversidad

Sobre GBIF.ES | Noticias y eventos | Datos de biodiversidad | Formación | Comunidad



Lanzamos el Plan Anual de Formación de GBIF España para 2025

La Unidad de Coordinación de GBIF España publica su PLAN ANUAL DE FORMACIÓN para 2025, compuesto por 8 talleres y 3 webinars. Siete...

Corvus monedulaoides - CC BY-NC 4.0 - [\[CC BY-NC 4.0\]](#)

Publicación

Publica tus datos y contribuye al conocimiento y conservación de la biodiversidad.

Uso de datos

Accede a millones de registros de biodiversidad a través de nuestros portales.

Gestión

Organiza y valida tus datos mediante nuestras herramientas informáticas.

Formación

Saca rendimiento a los datos a través de nuestros talleres de formación.



Mostrando 1 a 20 de 514

Logo	Nombre	Organización	Tipo	Subtipo	Registros	Última modificación	Última publicación
	Lista actualizada de los microlepidópteros paleárticos (Insecta, Lepidoptera) del Museo de Ciencias Naturales de Barcelona	Museo de Ciencias Naturales de Barcelona	Checklist		668	2025-01-14 14:42:38	2025-01-14 14:42:40
	Barakaldo: A database compiling taxonomic and distributional data on Ibero-Balearic arthropods	Departamento de Ecología y Hydrología, Universidad de Murcia	Documento		1,006,768	2025-01-14 13:12:17	2025-01-14 13:12:05
	Barakaldo: A database compiling taxonomic and distributional data on Ibero-Balearic arthropods	GBIF-Spain	Documento		1,006,768	2025-01-14 13:10:41	2025-01-14 13:10:16
	Aranzadi Ringing Scheme (bird ring recovery)	Aranzadi Science Society	Observación		479,240	2025-01-09 11:00:31	2025-01-09 11:00:57
	Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MAB)	Departamento de Biología, Ecología y Evolución de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MAB)	Documento		125,920	2025-01-08 09:44:04	2025-01-08 09:44:11
	Checklist of earthworm species (Oligochaeta) of the Andaman and Nicobar Islands, India	Museo de Ciencias Naturales de Barcelona	Checklist		28	2024-12-20 11:49:18	2024-12-20 12:26:43
	Flora vascular de	Universitat de les Illes Balears	Documento	Observación	143,163	2024-12-20	2024-12-20

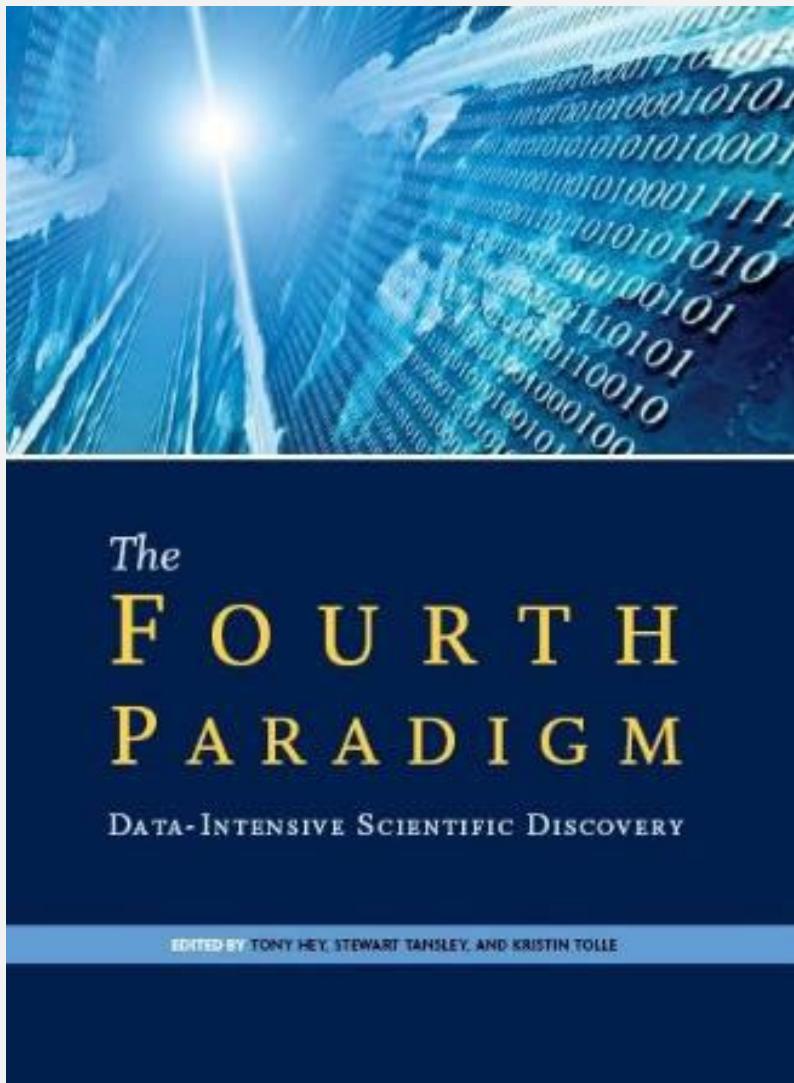
Most popular courses



Entidades del País Vasco que publican en GBIF

Institución / Entidad	Enlace a su página en GBIF	Nº registros
Basoinsa S.L.	https://www.gbif.org/publisher/4701cb61-2296-4d74-bc87-525303218515	1.026
Departamento Zoología y Biología Celular Animal, Universidad del País Vasco	https://www.gbif.org/publisher/0f797329-9a45-44fe-a3f5-358638ccf62b	2.509
Universidad del País Vasco. Laboratorio de Botánica. Dpto. de Biología Vegetal y Ecología	https://www.gbif.org/publisher/e8fb330c-5750-4698-afba-5438618af521	8.580
Gobierno Vasco	https://www.gbif.org/publisher/743cc32d-d83d-4e3a-a574-77d1725b1e08	191.383
Museo de Ciencias Naturales de Álava	https://www.gbif.org/publisher/4c91866b-3c2e-4568-aca0-3ab0a1c1a45e	246.379
Jardín Botánico de Olarizu (Vitoria-Gasteiz)/Olarizuko Lorategi Botanikoa (Vitoria-Gasteiz)	https://www.gbif.org/publisher/0e5e2ebb-cc35-4e96-ac4c-298654ff97d6	577
Sociedad de Ciencias Aranzadi	https://www.gbif.org/publisher/823818f6-0696-4e29-bc4c-b6f3817535a7	852.086
AZTI (colecciones biológicas marinas)	https://www.gbif.org/publisher/33867370-b176-4f56-81ed-f7e60be9668e	46.156
Sociedad Micológica Barakaldo	https://www.gbif.org/publisher/229f478b-fe5c-4470-abde-d960b59cea9a	4.623

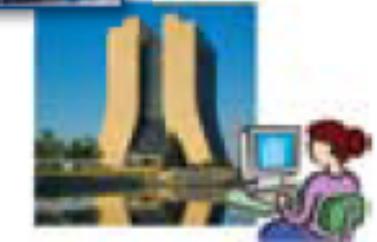
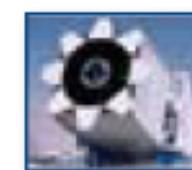
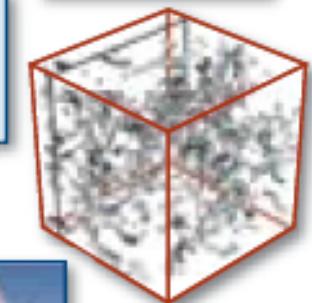
Ciencia y datos



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

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



Datos y conocimiento

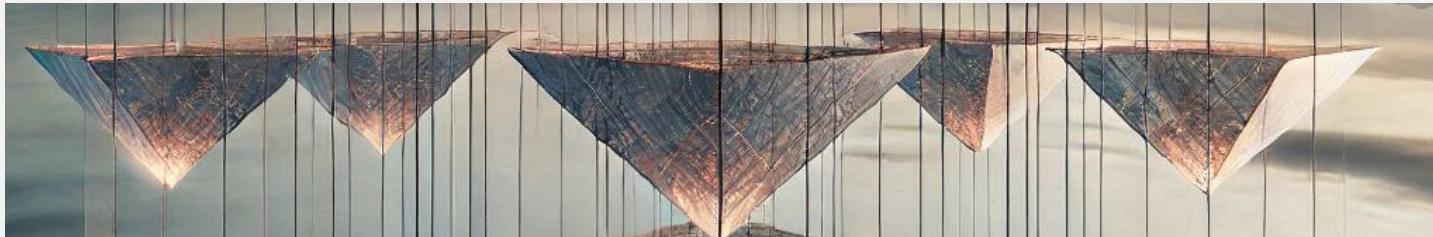
“
Data
is the
new
oil

Clive Humby,
cientista de datos



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

Henry Nix



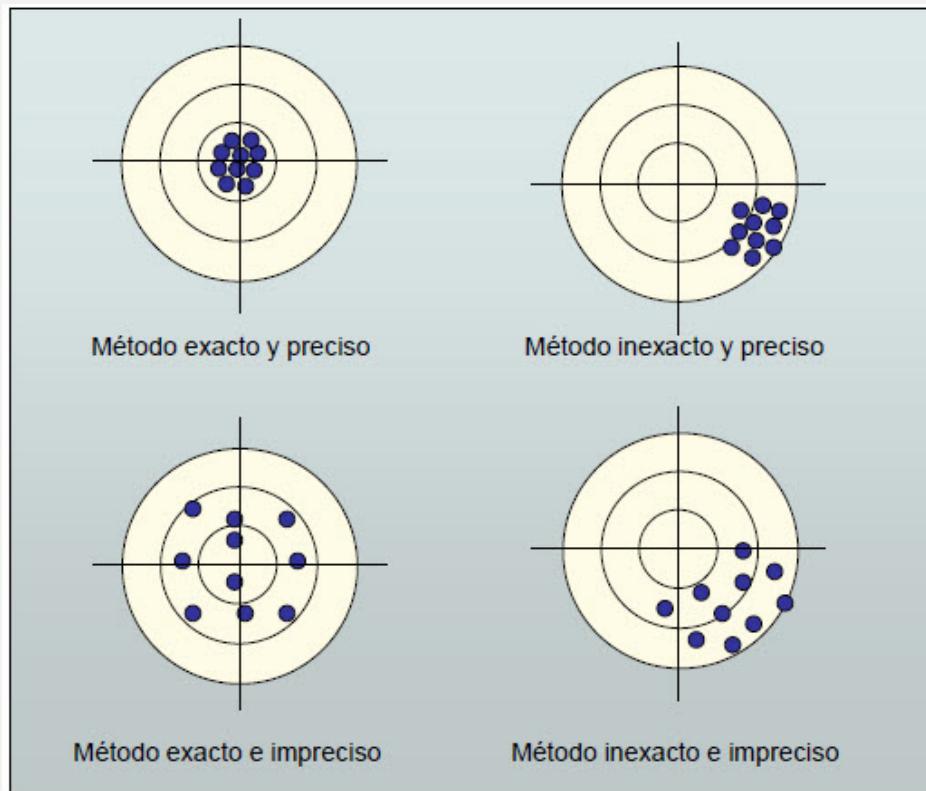
¿Que hace un dato bueno?



- Que sea “de calidad” frente a “que me sirva”
 - “fitness for use” -- adecuación para el uso - - erabilrerarako egokitasuna
- + Potencial de uso

Cambio de perspectiva, así vamos:
del dato bueno al dato útil (o potencialmente útil)
No hay dato malo

Que hace un dato útil. La parte intrínseca



- Que sea fidedigno;
 - Precisión y exactitud
 - Resolución
 - Completitud

Synced 3 years ago

OCCURRENCE | UNKNOWN DATE

Passiflora Glauca

Passionflower In English Recorded in Colombia

Plantae > Tracheophyta > Magnoliopsida > Malpighiales > Passifloraceae

DETAILS

GBIF Taxon interpretation: Passiflora L.

Dataset: Real Jardín Botánico: Dibujos de la Real Expedición Botánica del Nuevo Reino de ...

Location: Colombia

Publisher: Real Jardín Botánico (CSIC)

Basis of record: Occurrence

Issues: Taxon match higherrank

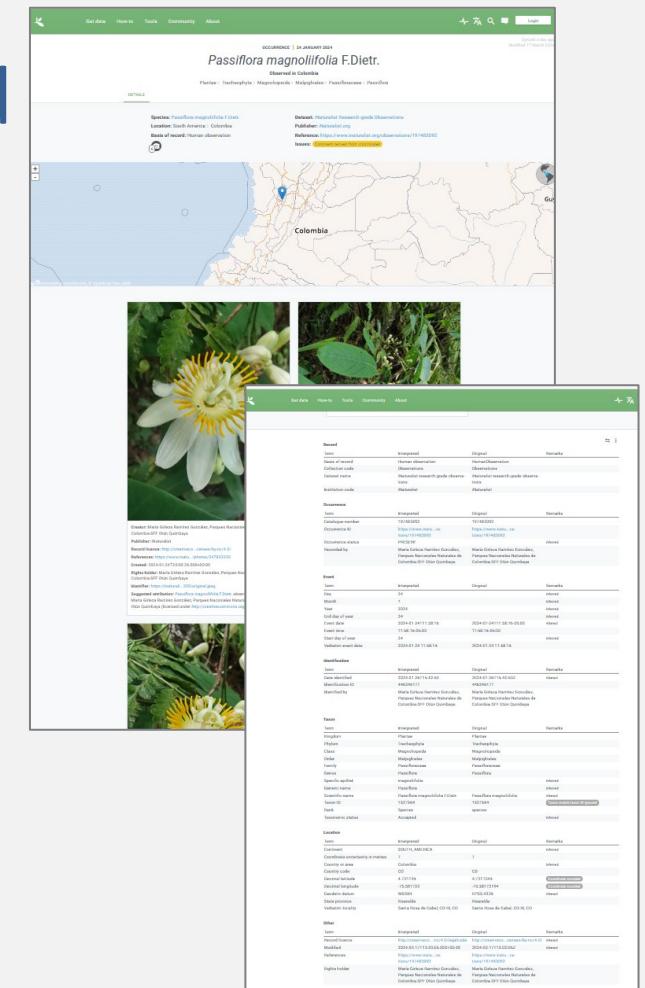
Coordinates missing

This record is published without coordinates, but it includes a textual description of its location.

Location: Colombia



<https://www.gbif.org/occurrence/420866219>



<https://www.quimicafisica.com/fundamentos-de-quimica/exactitud-y-precision.html>

<https://www.gbif.org/occurrence/439866319>

<https://www.gbif.org/occurrence/4516416057>

Que hace un dato útil. La parte del contexto (metadatos)

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 o posibles problemas con los datos?

8. Acceso y uso:

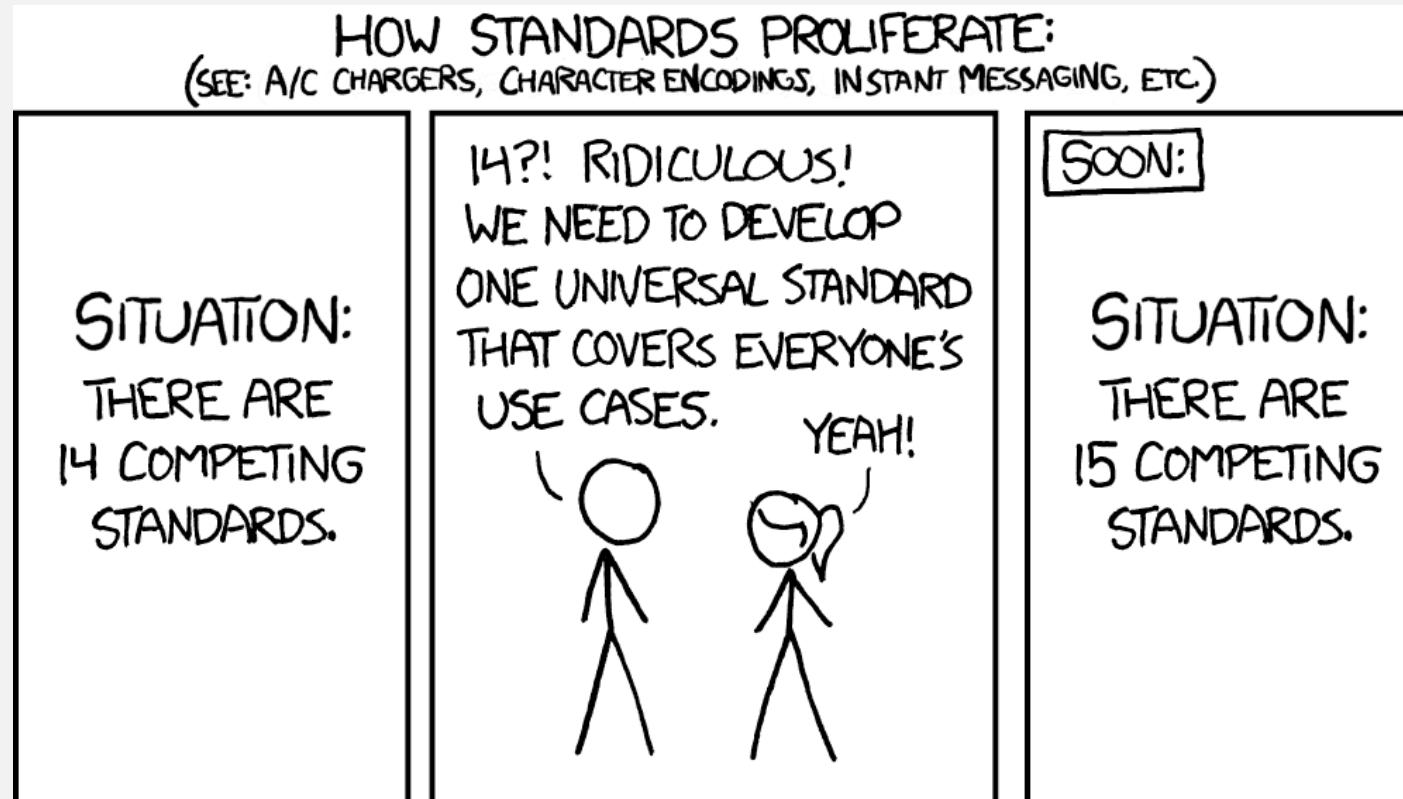
1. ¿Quién puede acceder a los datos?
2. ¿Existen restricciones en el uso de los datos?

9. Relaciones:

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

Estándares

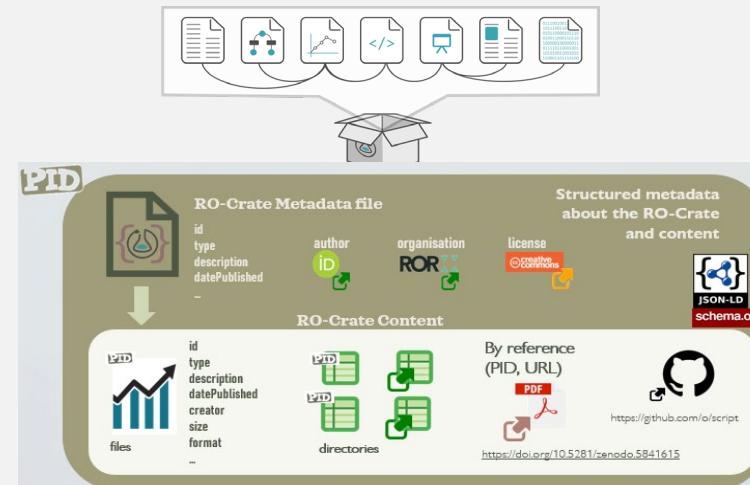
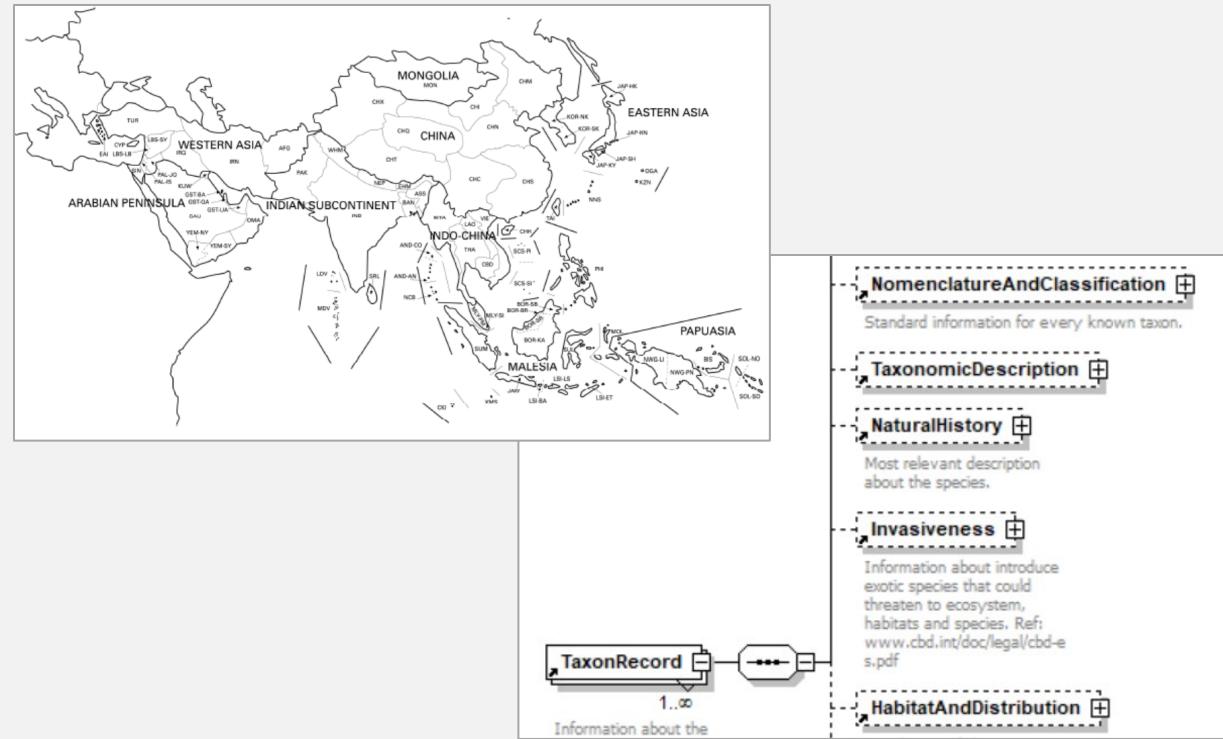
“pensaba que después del COVID no tendría que contar estas cosas...”



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

Estándares

- Vocabularios controlados (léxicos)
 - Que valores son válidos
 - (listas de géneros, de países...)
- Definiciones (semánticos)
 - para humanos, para máquinas
 - (v.gr.: Darwin Core [terms])
- Estructurales (sintácticos)
 - Como se codifica la información
 - (v.gr.: Frictionless Data Packages)



Licencias



LICENSES **TERMS**

Attribution
BY
Others can copy, distribute, display, perform and remix your work if they credit your name as requested by you

No Derivative Works
ND
Others can only copy, distribute, display or perform verbatim copies of your work

Share Alike
SA
Others can distribute your work only under a license identical to the one you have chosen for your work

Non-Commercial
NC
Others can copy, distribute, display, perform or remix your work but for non-commercial purposes only.

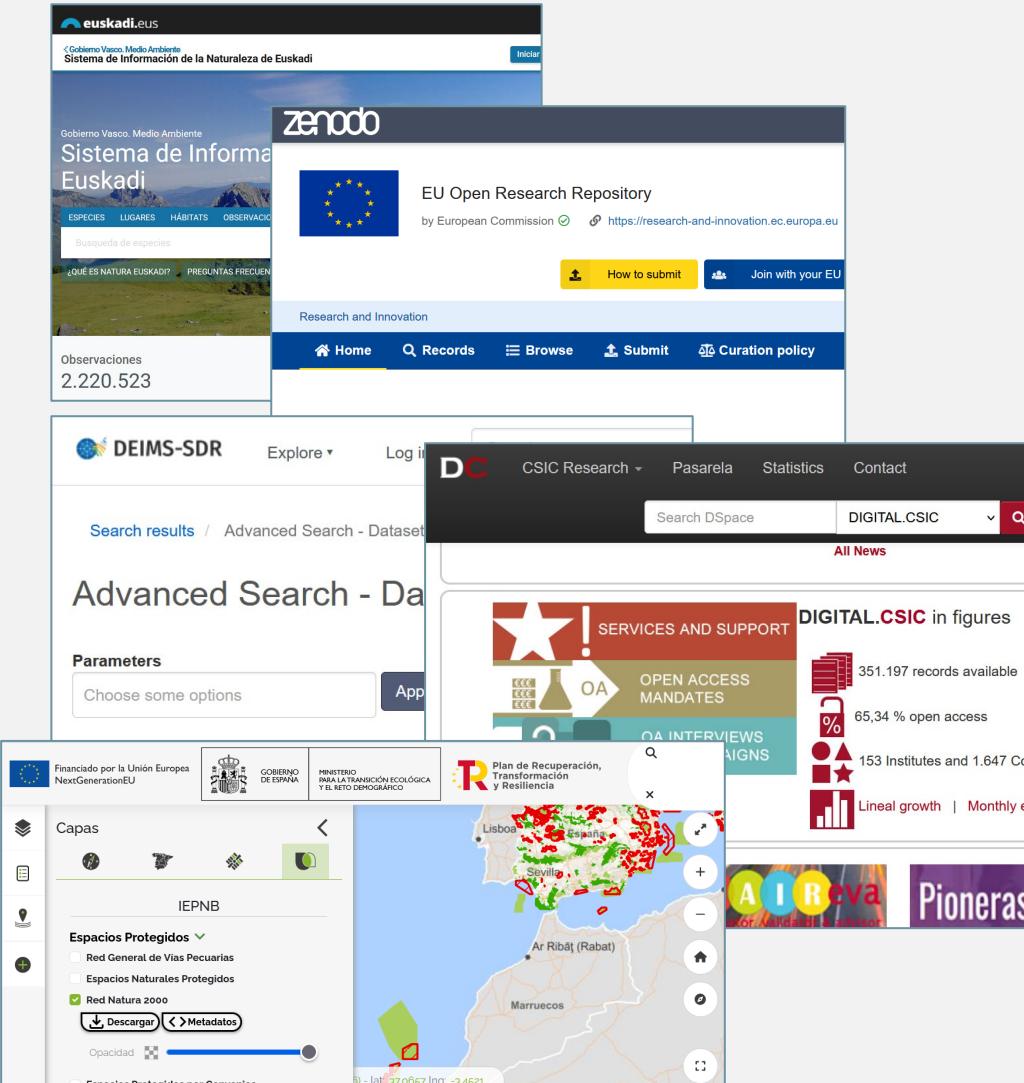
“Estándares orquestados”

- CC es un vocabulario controlado
- Dublin Core es un estándar de términos
- XML es un estándar sintáctico

Xml

```
<metadata>
  <dc:title>Example Document</dc:title>
  <dc:creator>John Doe</dc:creator>
  <dc:subject>Metadata Example</dc:subject>
  <dc:description>This is an example of a metadata section in XML.</dc:description>
  <dc:publisher>Example Publisher</dc:publisher>
  <dc:date>2025-01-23</dc:date>
  <dc:type>Text</dc:type>
  <dc:format>application/pdf</dc:format>
  <dc:identifier>http://example.com/document</dc:identifier>
  <dc:language>en</dc:language>
  <dc:rights>Creative Commons Attribution-NonCommercial 4.0 International</dc:rights>
</metadata>
```

Como se sirve ese dato

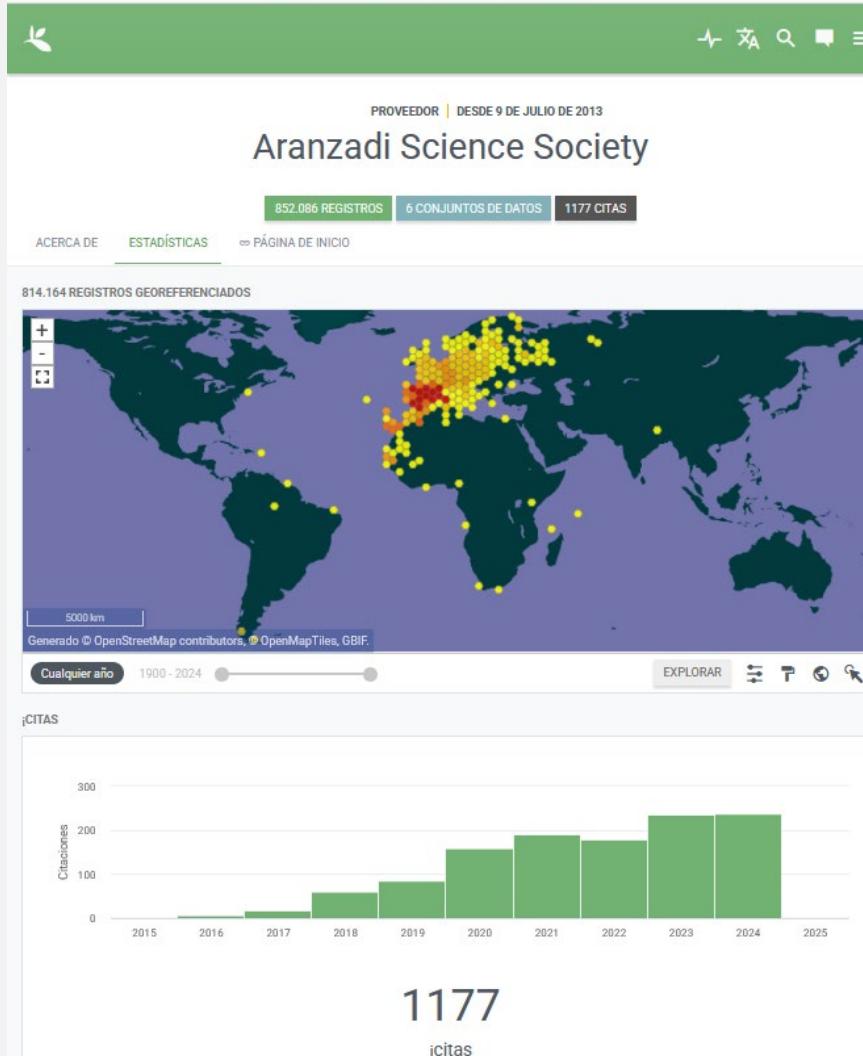


- Hablamos de las plataformas, portales, repositorios y sus funcionalidades
 - ¿Van los datos con sus metadatos?
 - Quien, donde cuando, como, para qué,...
 - ¿Están indexados?
 - ¿Se pueden descargar?
 - ¿Tienen licencia?
 - ¿Se usa estándares (para los datos y los metadatos)?
 - ¿Están identificados?

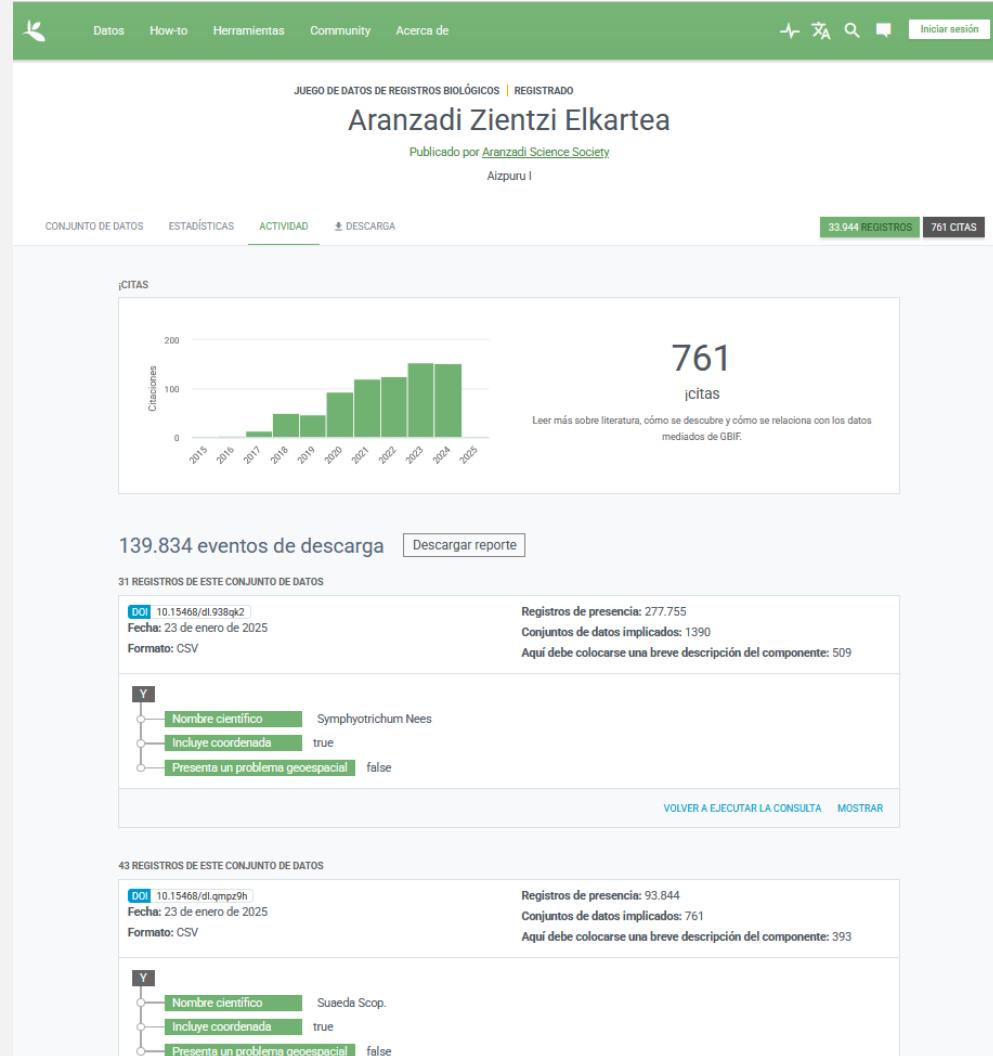
Identificadores

- Útiles para el usuario: transparencia, credibilidad, gestión, eficiencia.... Y por ello los venimos usando “desde siempre”: signaturas 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)
- Los CPR de Dinamarca
- `550e8400-e29b-41d4-a716-446655440000`

Identificadores: Para juegos de datos en GBIF, los DOIs



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



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

- todo lo anterior, y además métricas
- Utilidad para quien los usa; sostenibilidad para quien los genera

¿Que cambia los DOIs?

(Identificadores únicos, universales, persistentes y resolvibles)

BUSCAR RECURSOS | 289 RESULTADOS

TODOS LITERATURA

Ler 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 a widely distributed in North America fir and spruce defoliator, intercepted in 2016 in the Netherlands. Analysis of the literature data on the known distribut...

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 biodiver...

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.25hgq [DOI](#) 10.15468/dl.2cw1t1
[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.2tfnps

- Si publicar en GBIF (o en otros portales) tenía impacto, ahora se cuento, 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>

Entidades del País Vasco que publican en GBIF (2)

Institución / Entidad	Enlace a su página en GBIF	Nº registros	Citaciones en publicaciones
Basoinsa S.L.	https://www.gbif.org/publisher/4701cb61-2296-4d74-bc87-525303218515	1.026	358
Departamento Zoología y Biología Celular Animal, Universidad del País Vasco	https://www.gbif.org/publisher/0f797329-9a45-44fe-a3f5-358638ccf62b	2.509	86
Universidad del País Vasco. Laboratorio de Botánica. Dpto. de Biología Vegetal y Ecología	https://www.gbif.org/publisher/e8fb330c-5750-4698-afba-5438618af521	8.580	680
Gobierno Vasco	https://www.gbif.org/publisher/743cc32d-d83d-4e3a-a574-77d1725b1e08	191.383	380
Museo de Ciencias Naturales de Álava	https://www.gbif.org/publisher/4c91866b-3c2e-4568-aca0-3ab0a1c1a45e	246.379	1.105
Jardín Botánico de Olarizu (Vitoria-Gasteiz)/Olarizuko Lorategi Botanikoa (Vitoria-Gasteiz)	https://www.gbif.org/publisher/0e5e2ebc-cc35-4e96-ac4c-298654ff97d6	577	332
Sociedad de Ciencias Aranzadi	https://www.gbif.org/publisher/823818f6-0696-4e29-bc4c-b6f3817535a7	852.086	1.166
AZTI (colecciones biológicas marinas)	https://www.gbif.org/publisher/33867370-b176-4f56-81ed-f7e60be9668e	46.156	111
Sociedad Micológica Barakaldo	https://www.gbif.org/publisher/229f478b-fe5c-4470-abde-d960b59cea9a	4.623	75

Una colateral. Datos e indicadores



Ley de Goodhart. Dice: "Cuando una medida se convierte en un objetivo, deja de ser una buena medida".

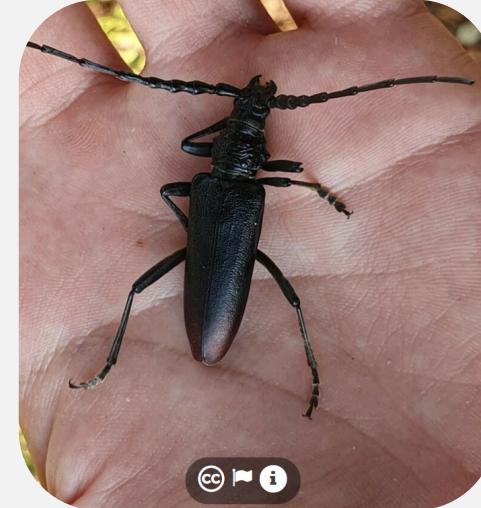
Esto significa que una vez que un indicador se utiliza como objetivo, a menudo pierde su efectividad como medida porque las personas comienzan a manipular el sistema para alcanzar el objetivo, en lugar de centrarse en el objetivo real

https://en.wikipedia.org/wiki/Goodhart%27s_law

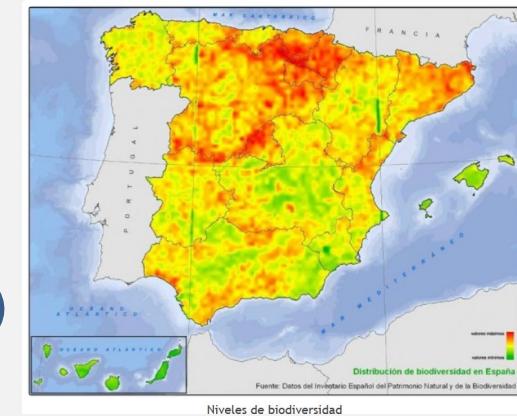
Otra colateral: Que es dato y que es información depende del punto de vista



Para el **taxónomo** los artejos de la pata del escarabajo son datos, la identificación es información



Para el **ecólogo** los puntos son datos, la distribución la información



Para el **técnico** las distribuciones son datos, las tendencias , la información

U2	U2	U2	U2	U2
XX	U2	U2	U2	U2
XX	U2	XX	U2	U2
FV	FV	FV	U1	U1
U1	U1	U1	U1	U1

Para la **política** los colores son los datos, el mensaje y las acciones, la información

The FAIR principles

Q 

FAIR Principles

Home > FAIR Principles

[Toggle menu](#)

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*.



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

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

Principios FAIR para los Datos Científicos



ENCONTRABLES



ACCESIBLES



INTEROPERABLES



REUTILIZABLES

... en GBIF

Registros de presencia de especies, datos de muestreos sistemáticos (seguimientos, monitoreos...), listas de especies

- Findable?
 - Indexed: data & metadata
 - Standardized: data & metadata
- Interoperable?
 - Inside network: full
 - Outside network: limited
 - SPARQL endpoint: no
- Accessible?
 - Cached: yes
 - Downloadable: yes
 - Visualizations: yes
- Reusable?
 - Downloadable: yes
 - Licenses schema: yes
 - DOIs for datasets & downloads
 - UUIDs for records: uneven implementation

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
- Documentar nuestros datos
- En vez de inventar nuestros estándares contribuir a la mejora de los que hay
- Citar las fuentes → DOIs
- No hay dato malo, pero si hay maneras de hacer los datos mejores

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://www.gbif.es/> , <https://datos.gbif.es>, <https://elearning.gbif.es>



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

