Gestión de datos de biodiversidad, en qué está la Ciencia ahora

Francisco Pando
(GBIF.ES-CSIC)

Gbif.es

GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA E INNOVACIÓN · CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Jornadas sobre Información de Biodiversidad y Administraciones Ambientales 2023

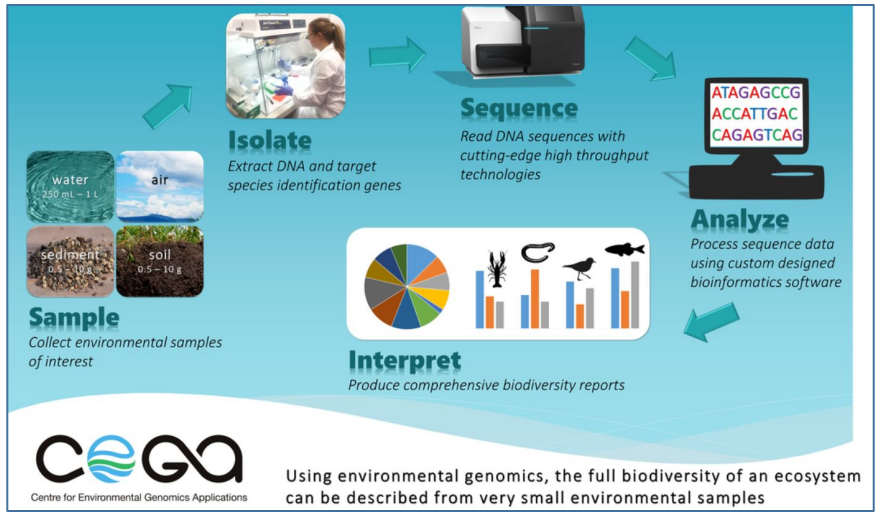
REAL JARDÍN BOTÁNICO Madrid
15-17 de noviembre de 2023

Sumario

- En episodios anteriores…
- El TDWG
- Sistemas de Información de Proyectos (AUS, MEX)
- Identificadores persistentes
- Wikidata
- “Digital Twins”
- Cubos de datos
- Una especificación de datos para muestreos sistemáticos
- Buenas prácticas y herramientas para georeferenciación retrospectiva
- “Frictionless data”
- “Workflows”

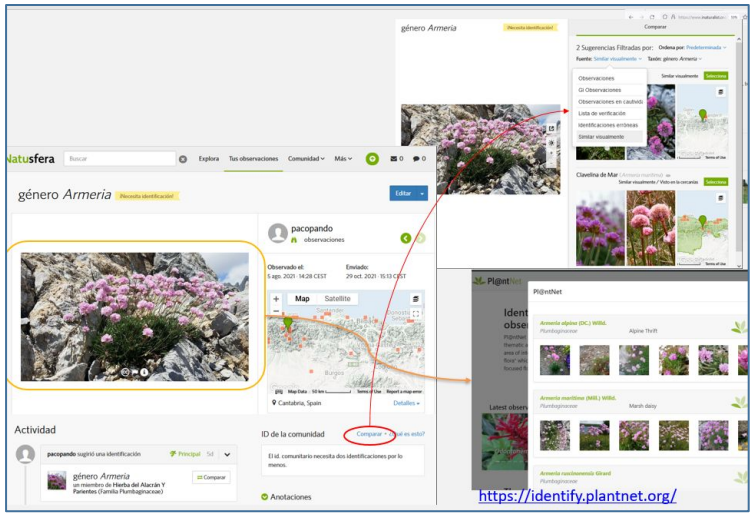
Como decíamos ayer…

Muestreos basados en ADN ambiental



https://www.cegacanada.com/about.html

Ciencia ciudadana e inteligencia artificial

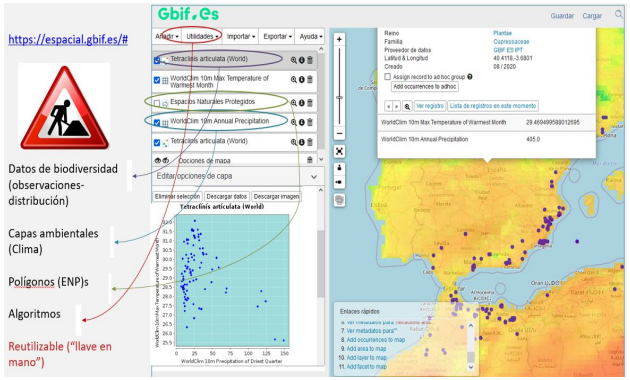


https://identify.plantnet.org/

World Flora Online

https://www.worldfloraonline.org/

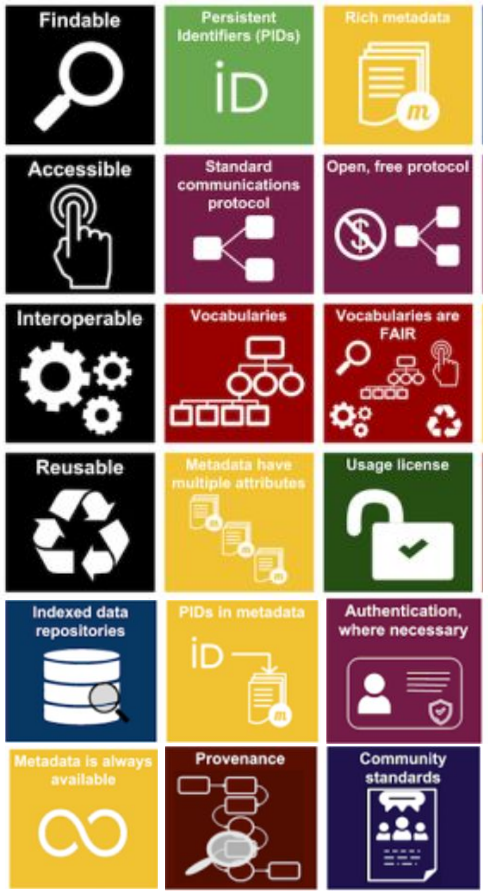
“VREs”



https://espacial.gbif.es/?lang=es

Como decíamos ayer…

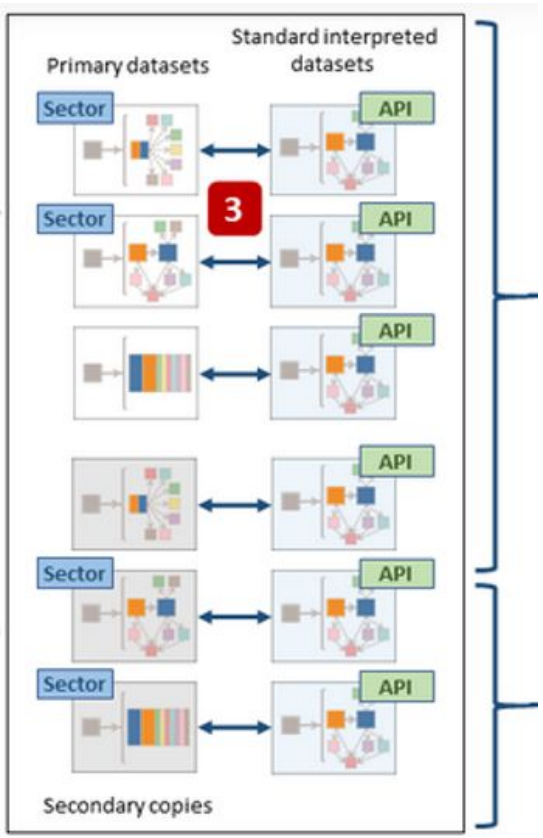
Principios FAIR



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

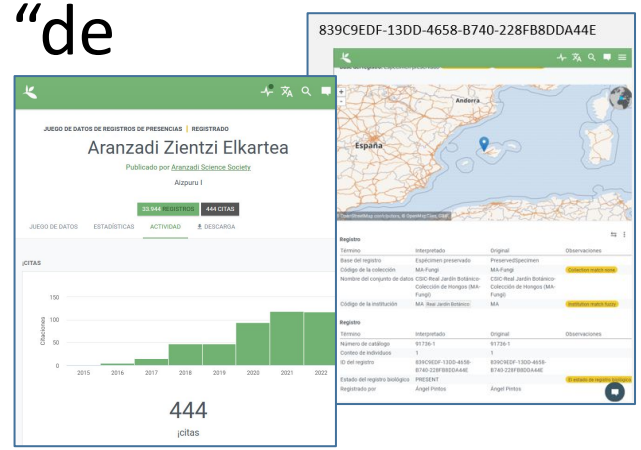
CheckListBank

Monitoring Evaluation Reporting and Improvement Tool



https://www.catalogueoflife.org/about/colpipeline

Identificadores “de largo alcance”



The new GBIF Data Model

MERIT

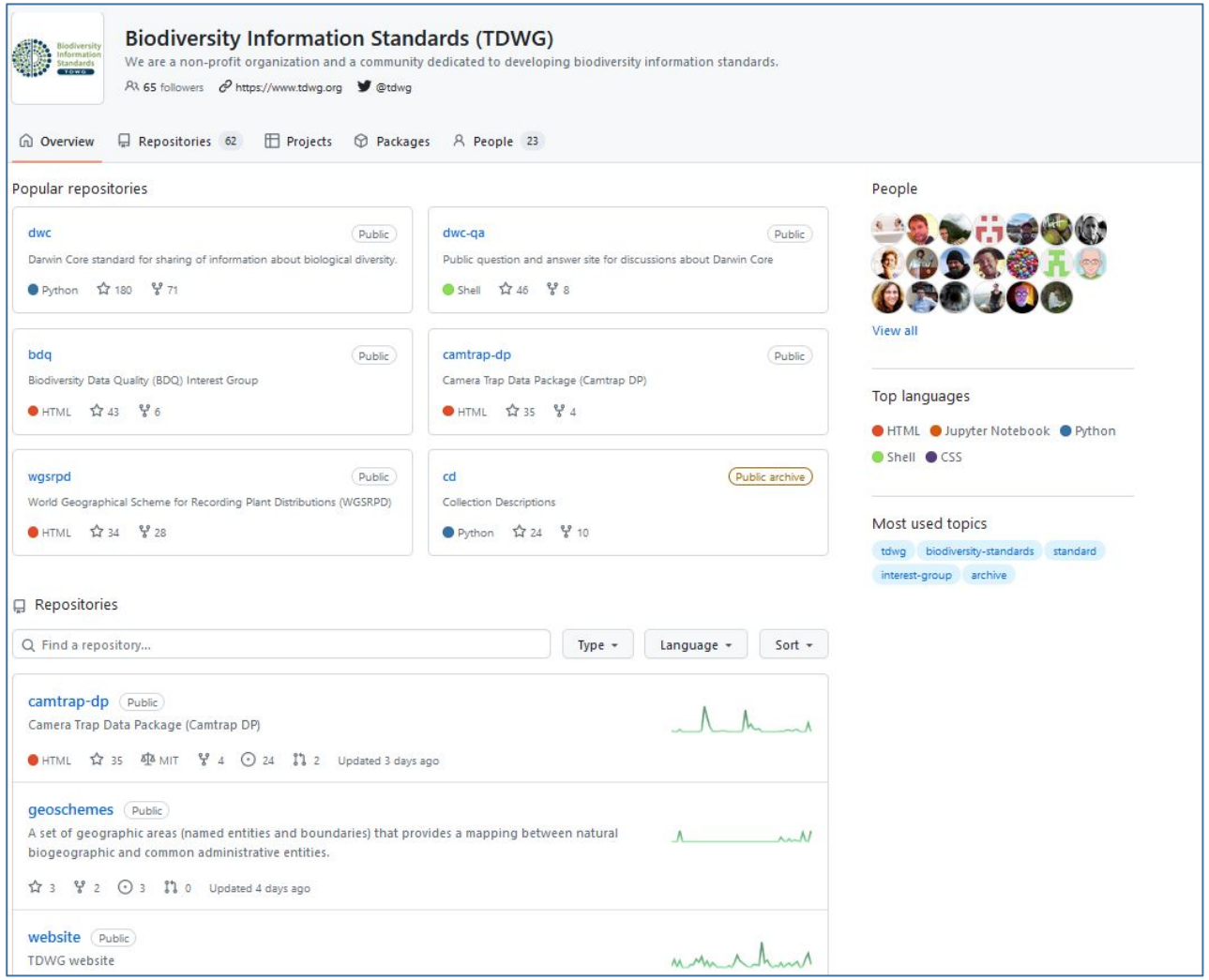
Monitoring, Evaluation, Reporting and Improvement (MERI) framework

TDWG: Biodiversity Information Standards

Biodiversity Information Standards
TDWG

- Access to Biological Collection Data (ABCD) Schema
- Audiovisual Core Multimedia Resources Metadata Schema
- Authors of Plant Names
- Botanico-Periodicum-Huntianum
- Botanico-Periodicum-Huntianum/Supplementum
- Darwin Core
- Description Language for Taxonomy (DELTA)
- Economic Botany Data Collection Standard
- Floristic Regions of the World
- GUID and Life Sciences Identifiers Applicability Statements
- Herbarium Information Standards and Protocols for Interchange of Data (HISPID3)
- Index Herbariorum. Part I: The Herbaria of the World
- International Transfer Format for Botanic Garden Plant Records (ITF2)
- Plant Names in Botanical Databases
- Plant Occurrence and Status Scheme (POSS)

https://www.tdwg.org/standards/

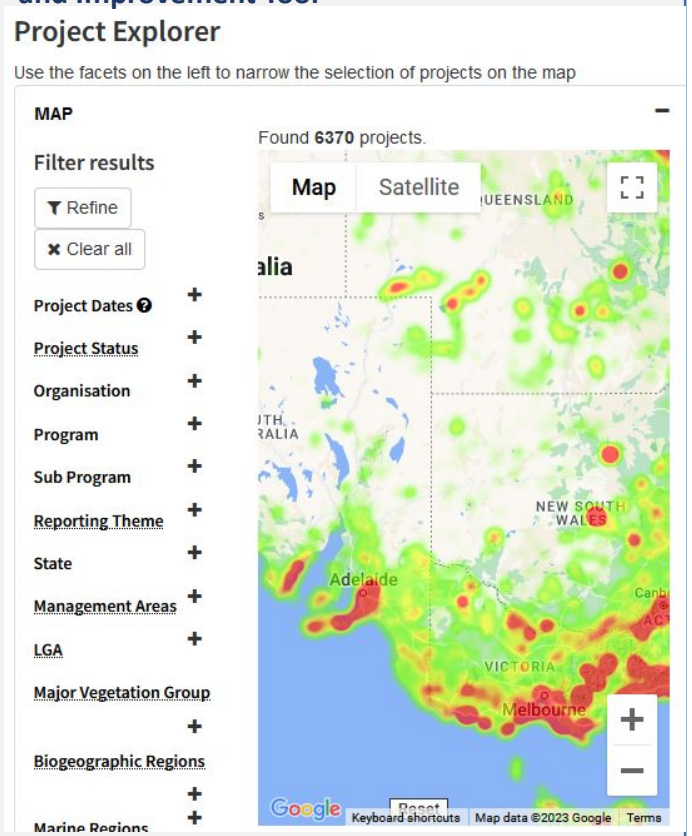


https://github.com/tdwg

Registro y seguimiento de proyectos

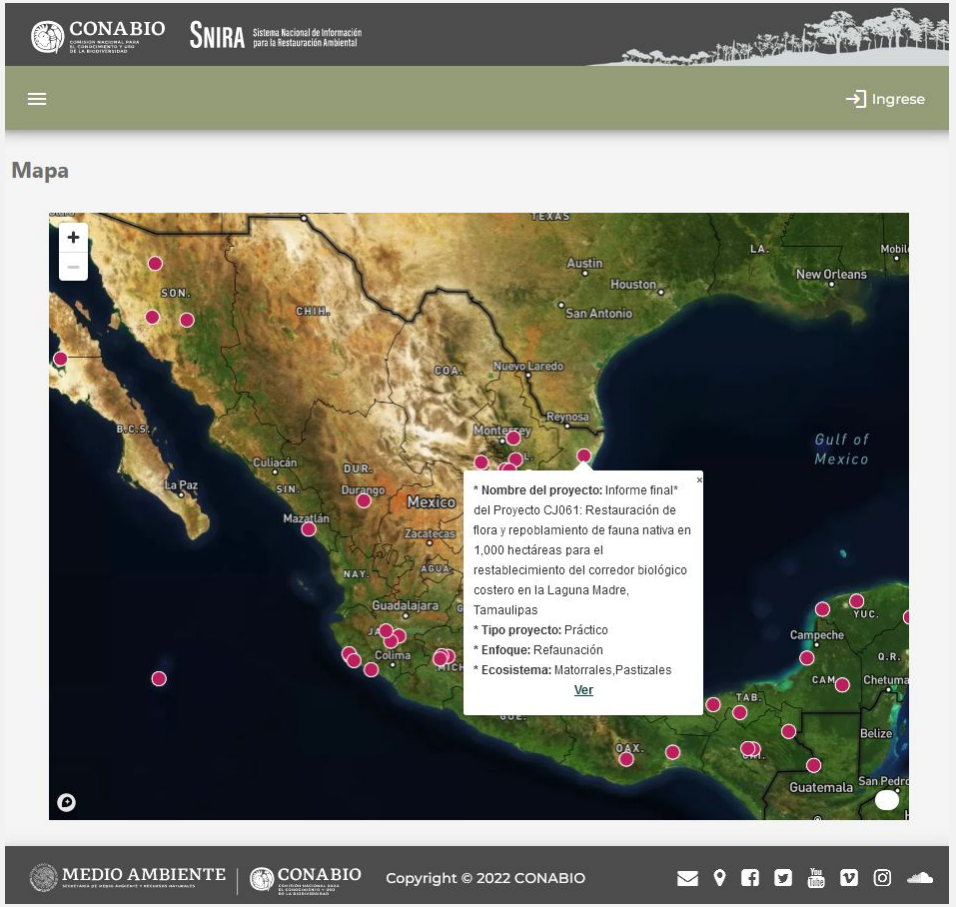
MERIT

Monitoring Evaluation Reporting and Improvement Tool



https://fieldcapture.ala.org.au/

Sistema Nacional de Información para la Restauración Ambiental (SNIRA)



https://snira.conabio.gob.mx/datos/mapa

Recovering the Kangaroo Island Narrow Leaved Mallee Woodland Threatened Ecological Community through local landscape action

☆ Add to favourites

Overview | Documents

Program	National Landcare Programme - Regional Land Partnerships	Project start	27-09-2018
Management Unit	Kangaroo Island	Project end	30-06-2023
Service Provider	KANGAROO ISLAND LANDSCAPE BOARD		
MERIT Project ID	RLP-MU17-P1		
Project status	ACTIVE		

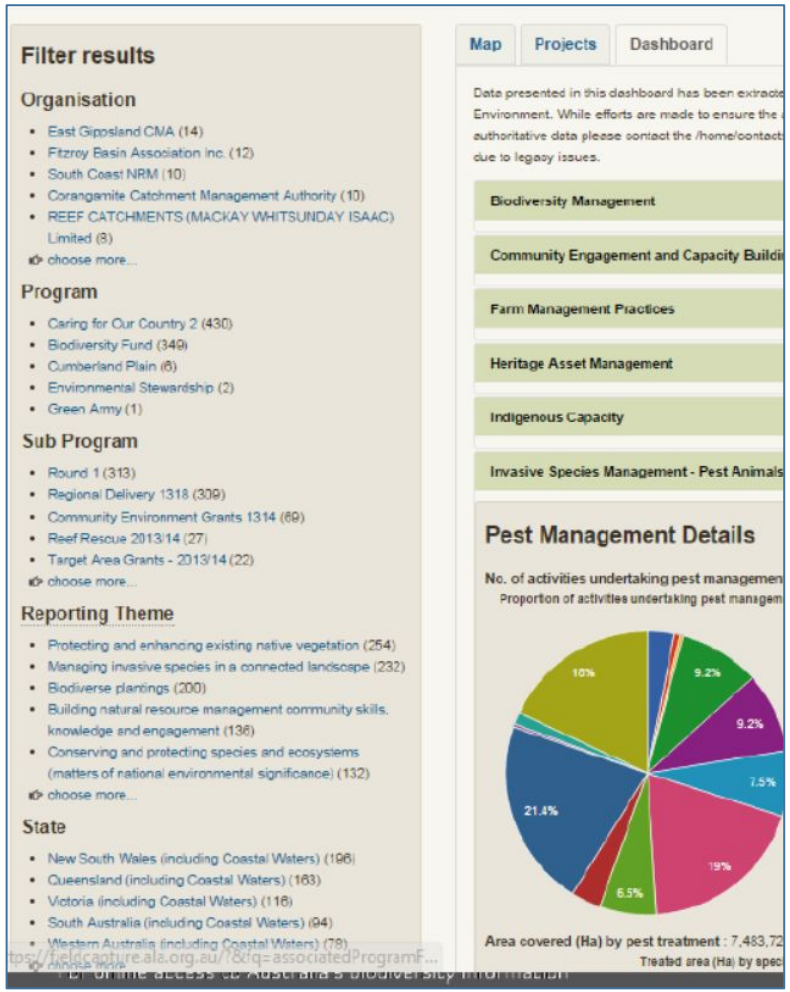
Program outcomes addressed

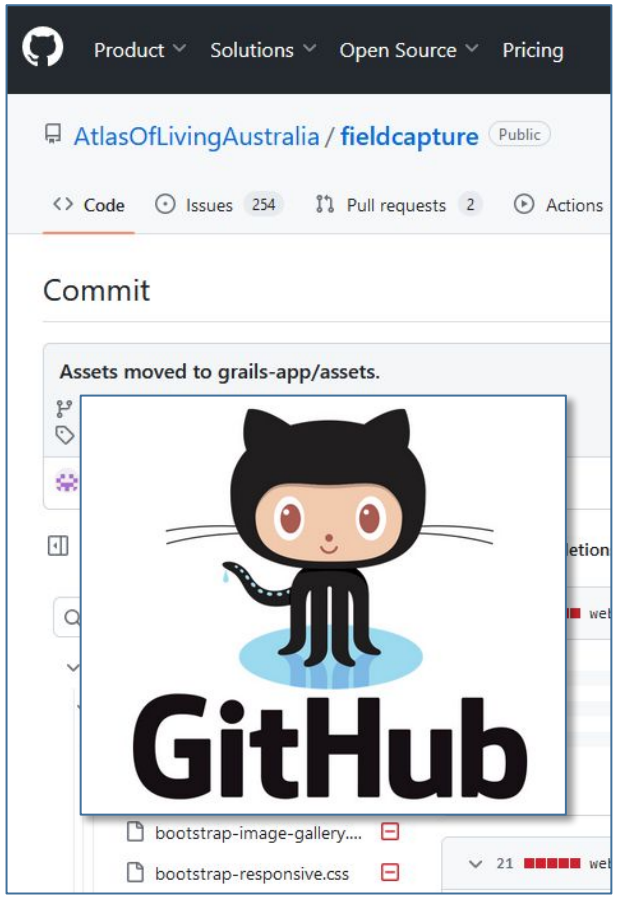
	Outcomes	Primary Investment Priority
Primary Outcome	4. By 2023, the implementation of priority actions is leading to an improvement in the condition of EPBC Act listed Threatened Ecological Communities.	Kangaroo Island Narrow-leaved Mallee (Eucalyptus cneorifolia) Woodland
Secondary Outcomes	2. By 2023, the trajectory of species targeted under the Threatened Species Strategy, and other EPBC Act priority species, is stabilised or improved.	Olearia microdisca (Small-flowered Daisy-bush)
	2. By 2023, the trajectory of species targeted under the Threatened Species Strategy, and other EPBC Act priority species, is stabilised or improved.	Leionema equestre (Kangaroo Island Phebalium)
	5. By 2023, there is an increase in the awareness and adoption of land management practices that improve and protect the condition of soil, biodiversity and vegetation.	Native vegetation and biodiversity on-farm

https://fieldcapture.ala.org.au/project/index/d31e6101-2d60-4351-b1ee-64f689f1d48f

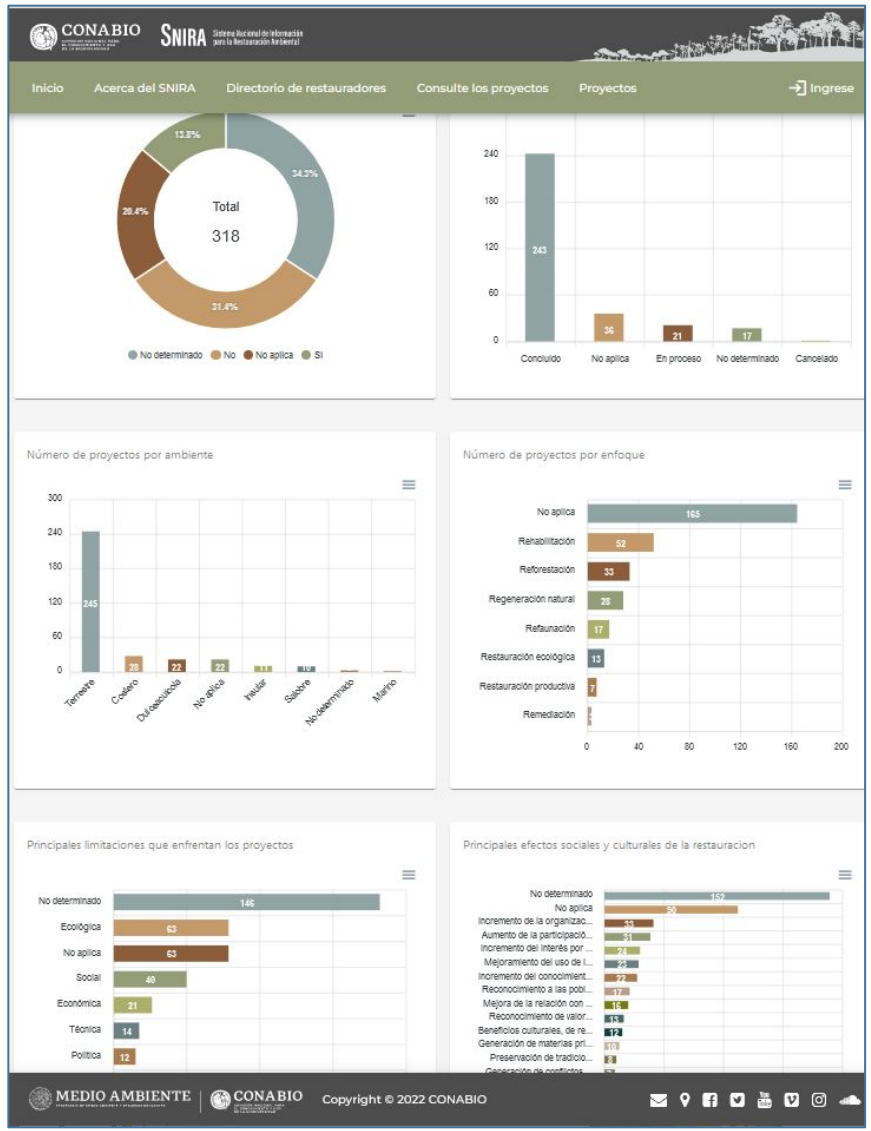
Registro y seguimiento de proyectos

- Aspectos destacables





https://github.com/AtlasOfLivingAustralia/fieldcapture



https://snira.conabio.gob.mx/datos/sintesis

Identificadores persistentes PIDs

Identificadores para:

referirnos a las cosas

agregar información

dar crédito – tener el context

Conectar contenidos, no páginas web

… en un entorno digital, distribuído, masivo ⇒

⇒ trazabilidad: del uso a la fuente, y de la fuente al uso

Multiples sistemas cada uno sus propias ventajas e inconvenientes. El uso de uno en concreto depende de factores como persistencia, la resolución, la interoperabilidad, la gobernabilidad, el coste y la adopción por parte de la comunidad

Identificadores persistentes PIDs

Algunos sistemas de PIDs en boga:

ORCID ([Identificador Abierto de Investigador y Colaborador]):autores con nombres similares, y a vincular sus publicaciones y afiliaciones en diferentes plataformas y bases de datos. ORCID es gratuito y abierto para cualquier persona que participe en actividades de investigación, erudición o innovación.
ROR ([Registro de Organizaciones de Investigación]): Este es un sistema de PI para organizaciones de investigación, como universidades, institutos, empresas, etc.
ARK ([Clave de Recurso de Archivo]): Este es un sistema de PI para cualquier tipo de objeto digital o físico, como páginas web, imágenes, vídeos, manuscritos, especímenes, etc. Los ARK están diseñados para ser flexibles, transparentes y de bajo coste, y son utilizados por muchas bibliotecas, archivos y museo.

Algunos que están ahí, pero sin mucha vigencia para proyectos nuevos:

LSID (Life Science Identifier): son identificadores persistentes y únicos para objetos digitales, como datos biológicos, documentos o imágenes.
PURLs (URL permanentes): actualmente administrado por Internet Archive. Tiene una base grande entidades que lo usan

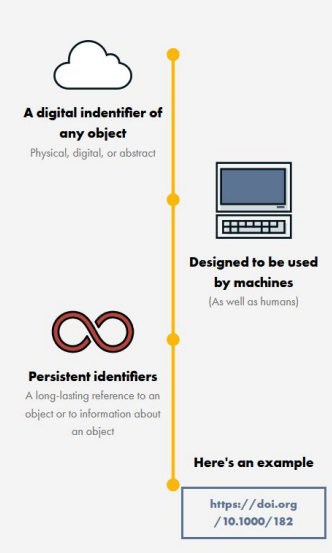
…Y los DOIs

Identificadores persistentes: PIDs

doi Foundation

The Identifier

WHAT IS A DOI?



A DOI is a digital identifier of an object, any object — physical, digital, or abstract. DOIs solve a common problem: keeping track of things. Things can be matter, material, content, or activities.

A DOI is a unique number made up of a prefix and a suffix separated by a forward slash. This is an example of one: 10.1000/182. It is resolvable using our proxy server by displaying it as a link: https://doi.org/10.1000/182.

Designed to be used by humans as well as machines, DOIs identify objects persistently. They allow things to be uniquely identified and accessed reliably. You know what you have, where it is, and others can track it too.

Despega con su uso para publicaciones

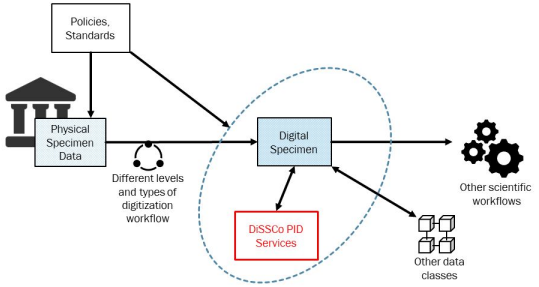
GBIF lo adoptó para juegos de datos y descargas

Uso en expansion a otras áreas

Resolver DOIs es gratis
Acuñar DOIs cuesta

Human & Digital (HAND) is , assigning DOI names to legal entity humans, licensed virtual humans, and fictional characters involved in the performing arts and sports.

https://digital.csic.es/



Digitally transforming collections science with Digital Specimens and persistent identifiers (PID). https://doi.org/10.3897/rio.7.e67379.figure1

Identificadores persistentes: PIDs

“Some communities may not be aware of the benefits and features of the DOI system, or may face barriers to access or use the system”

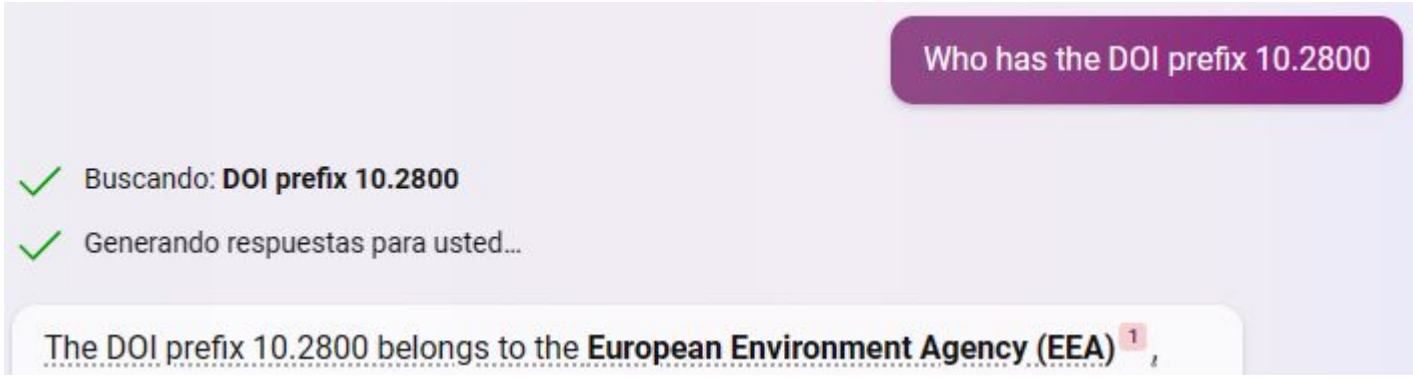
The European environment

State and outlook 2020 : knowledge for transition to a sustainable Europe

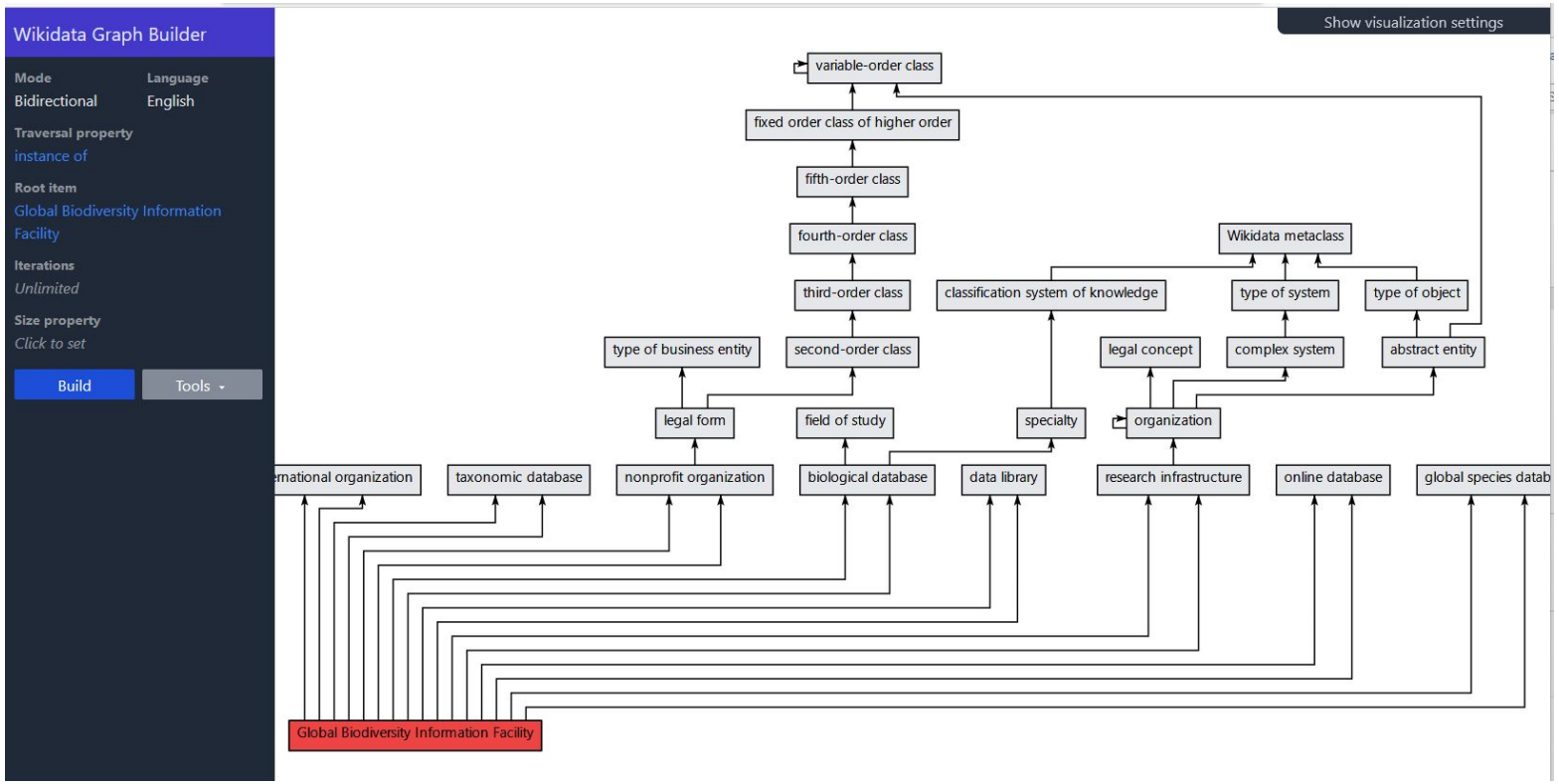
In 2020, Europe faces environmental challenges of unprecedented scale and urgency. Although EU environment and climate policies have delivered substantial benefits over recent decades, Europe faces persistent problems in areas such as biodiversity loss, resource use, climate change impacts and

View more

http://doi.org/10.2800/488115



Wikidata



https://angryloki.github.io/wikidata-graph-builder/?item=Q1531570&property=P31&mode=both&graph_direction=up

data blog

posts community-forum gbif.org about

Does Biodiversity Informatics 💘 Wikidata?

Quentin Groom & Deborah Paul
2020-02-21 · GBIF · Guest Post

Open online APIs are fantastic! You can use someone else's infrastructure to create workflows, do research and create products without giving anything in return, except acknowledgement. But wait a minute! Why is everyone not using them? Why do we create our own data sources and suck up the costs in time and money? Not to mention the duplication of effort.

One of the reasons is that we can't fix someone else's data. If we're lucky we can email them a list of the errors and hope for a response, but this is not a viable solution for a process that we would hope to be automated. So what if we all go local? Then we don't benefit from the latest data and the curation of experts in their domain and we all have to repeat the same work everywhere that the data are needed, particularly for people names, locality data, taxonomic names and concepts.

Is there a third way?

What about Wikidata? It's open, readable and writable to machine and human, curated by an avid and active group of people globally and, most importantly if you find errors you can fix them.

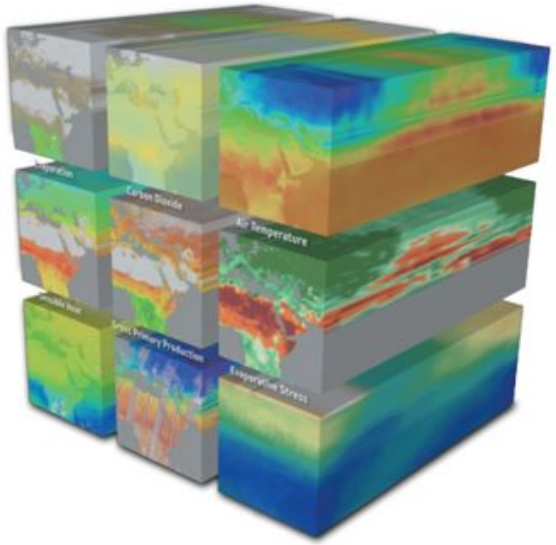
https://data-blog.gbif.org/post/wikidata/

Cubos de datos y biodiversidad

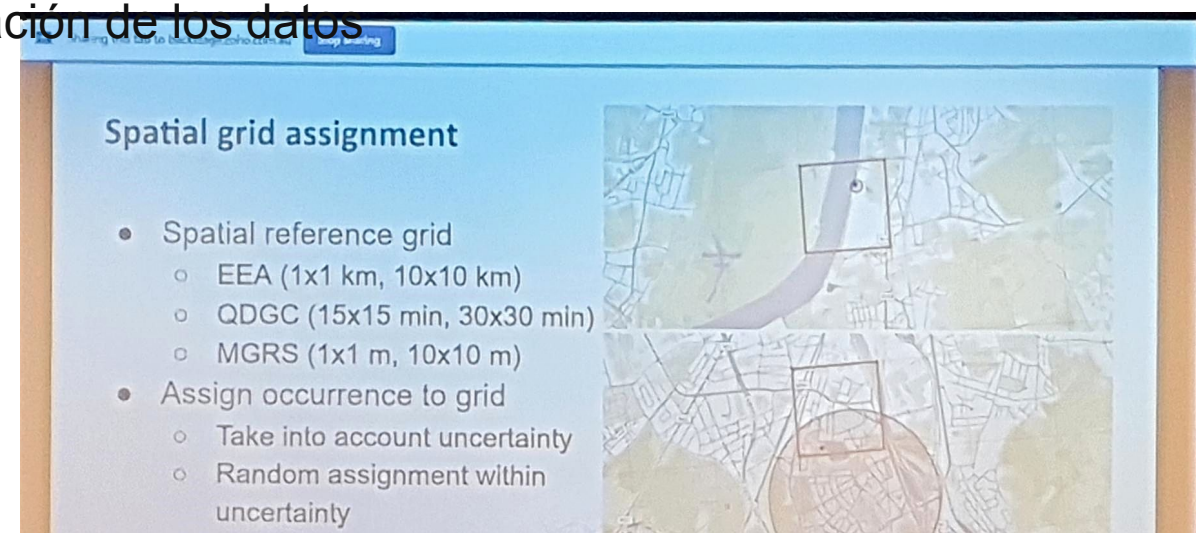
Un cubo de datos es una forma de representar los datos en múltiples dimensiones. La unidad es la celda, cada celda puede tener múltiples dimensiones. Es un formato orientado al análisis

BIODIVERSITY BUILDING BLOCKS FOR POLICY

B-Cubed es un proyecto europeo que tiene como objetivo desarrollar una plataforma de datos de biodiversidad basada en la nube, que integre diferentes fuentes de información y que facilite el análisis y la visualización de los datos

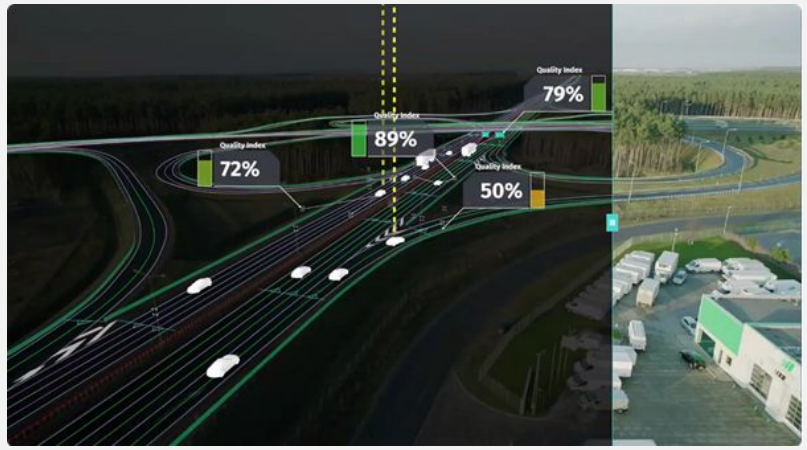


https://eo4society.esa.int/2019/05/21/european-data-cube-facility-service-an-eo-resource-factory/



https://b-cubed.eu/

“Digital Twins”

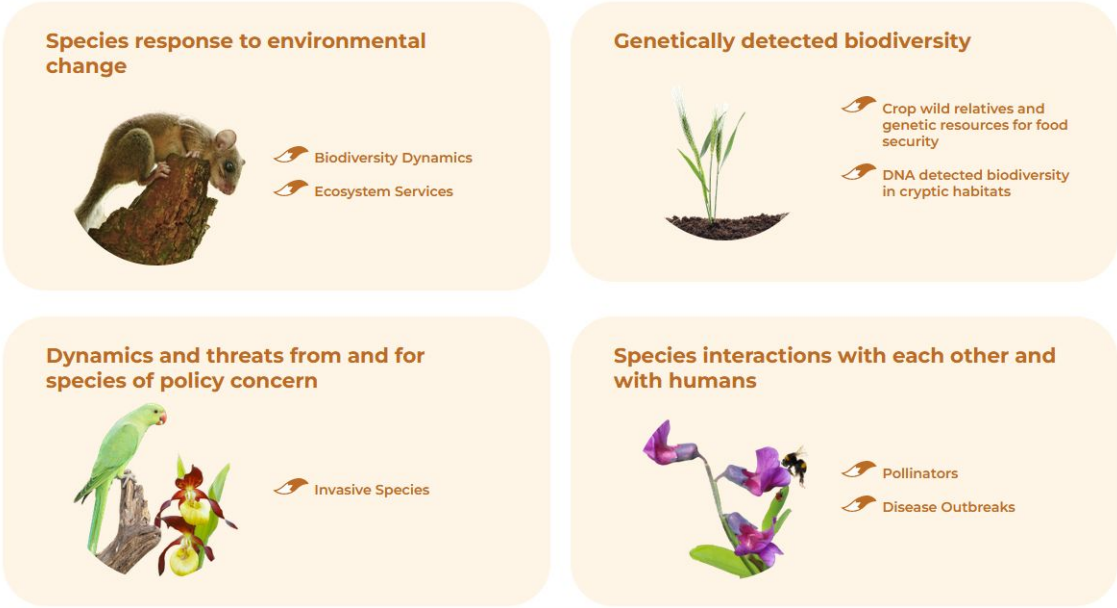


- Representación virtuales de un objeto o sistema físico que se actualiza con datos en tiempo real y utiliza la inteligencia artificial y el aprendizaje automático para simular su comportamiento y rendimiento en la vida real.
- Son representaciones de propósito específico.

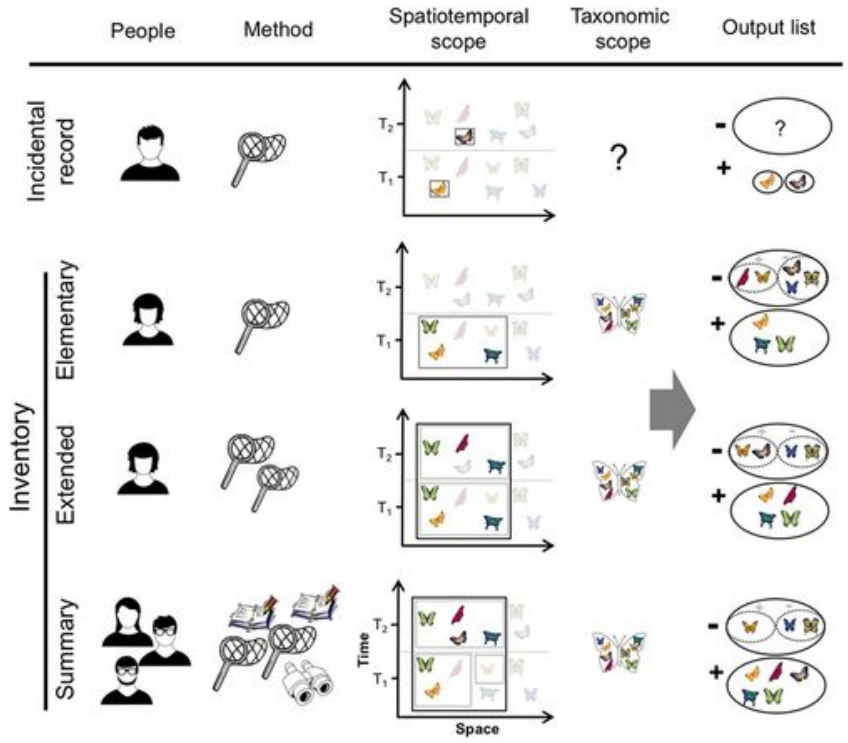
← https://www.arrb.com.au/news/digital-twin-technology-and-the-roads-sector

BIODT
biodiversitydigitaltwin

https://biodt.eu/



Una especificación de datos para muestreos sistemáticos



Humboldt Core Version 1	Single Sourced Inventories (e.g., transect count, trapping and netting, gridded atlas survey)	Summary Inventories (e.g., CTFS forest, relevé plots, protected area species lists)
General Dataset & Identification Terms	inventory performed by; dataset name, identifier, publisher, license, rights holders; metadata recorded by; citation reference and id; taxa identified by; identification quality; cited taxonomic authority	
Geospatial & Habitat Scope Terms	geospatial scope; areal extent; total area inventoried; number of sites; site names and details; lat/long by site; elevation range and units; habitats included and excluded	
Temporal Scope Terms	survey time blocks; start and end year, month day; time units spent In blocks; daily start, end time; study diurnality, study season	
Taxonomic Scope Terms	prospective taxonomic scope inclusion and exclusion; distribution status included and excluded; developmental stage included and excluded, size classes included and excluded	
Methodology Description Terms	inventory type; protocol name, detail, citation; reference; abundances reported Y/N; absences reported Y/N?	inventory type; compiled data Y/N, type; abundances and/or absences reported?; absence list
Completeness & Effort Terms	effort reporting, lower/upper bound, granular breakdown; effort method; vouchers/samples taken, if so, how? completeness reporting if necessary	completeness reported, how assessed; inferred taxonomic completeness upper/lower bound, how assessed

Public review of Humboldt Extension to Darwin Core

September 07, 2023

The public review and comment period for terms comprising the Humboldt Extension is now open for a minimum of 30...

https://www.tdwg.org/news/2023/humboldt-extension-public-review/

https://doi.org/10.1111/ecog.02942

https://www.tdwg.org/community/osr/humboldt-extension/

Buenas prácticas y herramientas para georeferenciación retrospectiva

GBIF Global Biodiversity Information Facility

Otros Formatos

PDF format
English

Para contribuciciones

Mejora este documento
Crear una propuesta
Editar en GitHub

Buscar

Tabla de Contenido

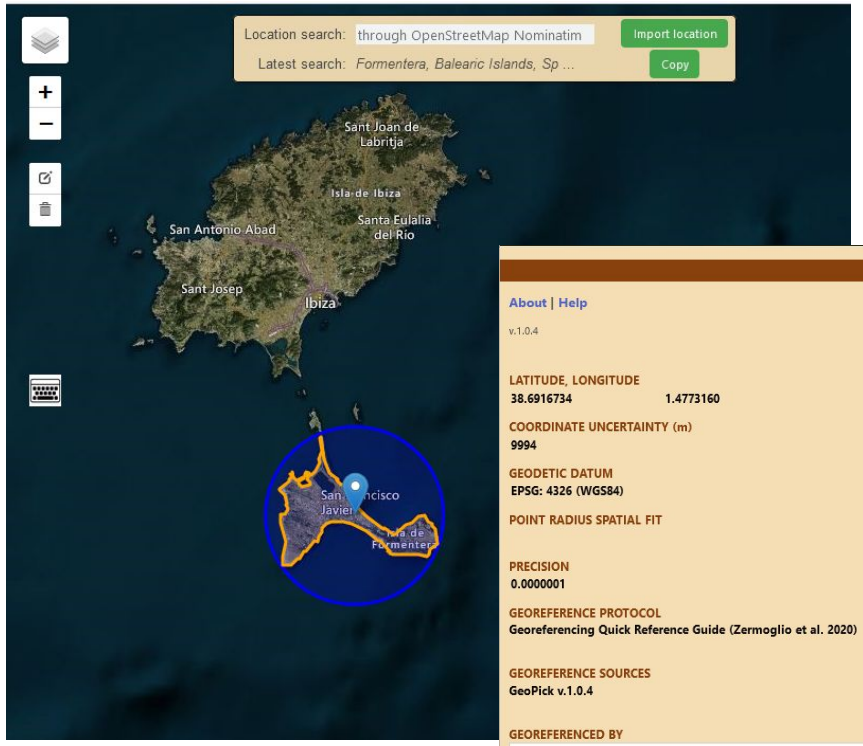
Guía rápida de georreferenciación

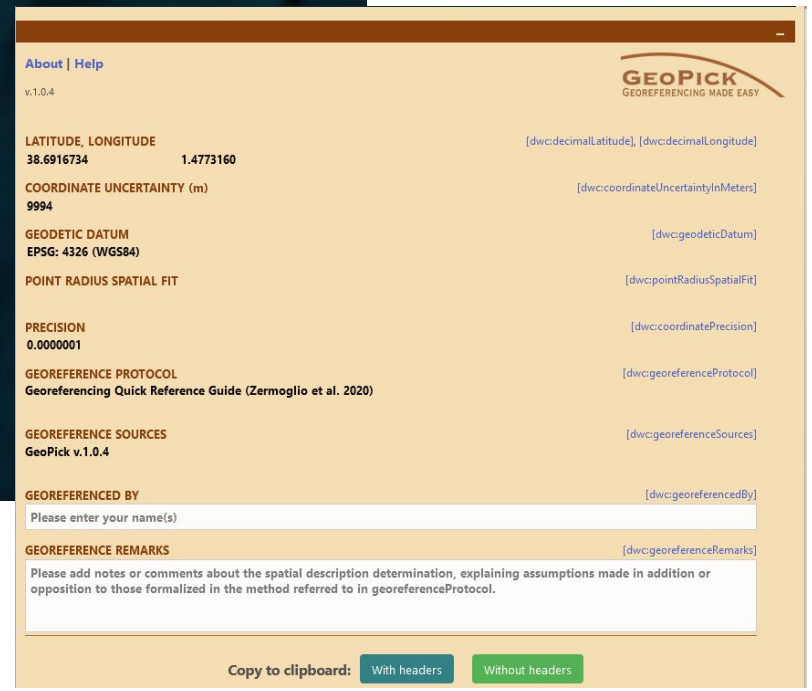
Paula F. Zermoglio · Arthur D. Chapman · John R. Wieczorek · María Celeste Luna · David A. Bloom – Versión 4ac9d96, 2022-02-25 15:09:32 UTC

Este documento también está disponible en formato PDF y en otros idiomas: English.

Colofón

Citación sugerida





https://doi.org/10.35035/e09p-h128

https://geopick.gbif.org/

Frictionless Data

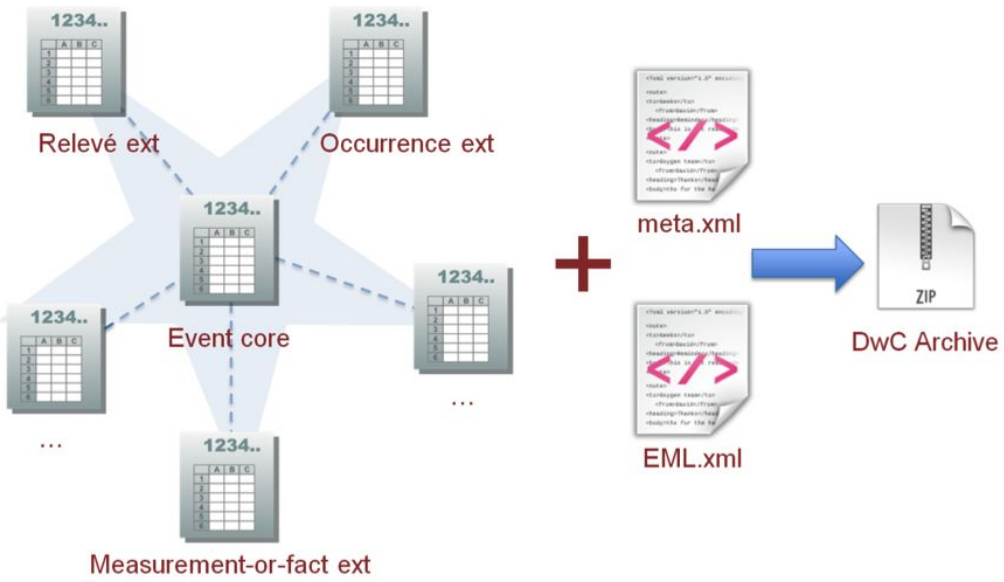
CC-By https://www.vecteezy.com/members/carterart

- A collection of standards and tools for publishing data. Supported by the Open Knowledge Foundation
- **Frictionless data** is about removing the friction in working with data through the creation of tools, standards, and best practices for publishing data using the Data Package standard, a containerization format for any kind of data. It offers specifications and software around data publication, transport and consumption. data resources are presented in CSV files while the data model is described in a JSON structure.

Being **domain agnostic**, Frictionless Data is open to all sorts of data, not just biodiversity related fields. It allows **truly relational** model as you can express primary and foreign keys in each of your data resources

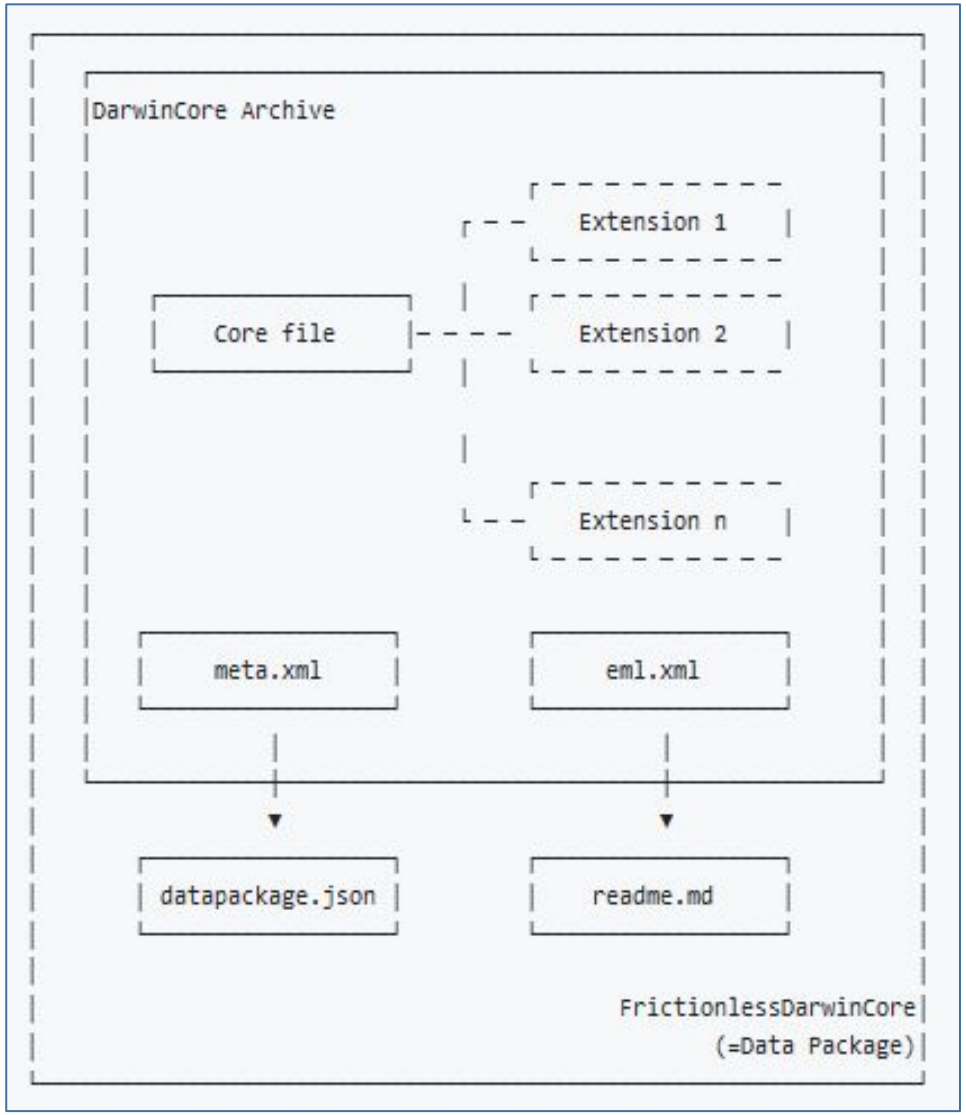
https://frictionlessdata.io/

“Frictionless data” para datos de biodiversidad



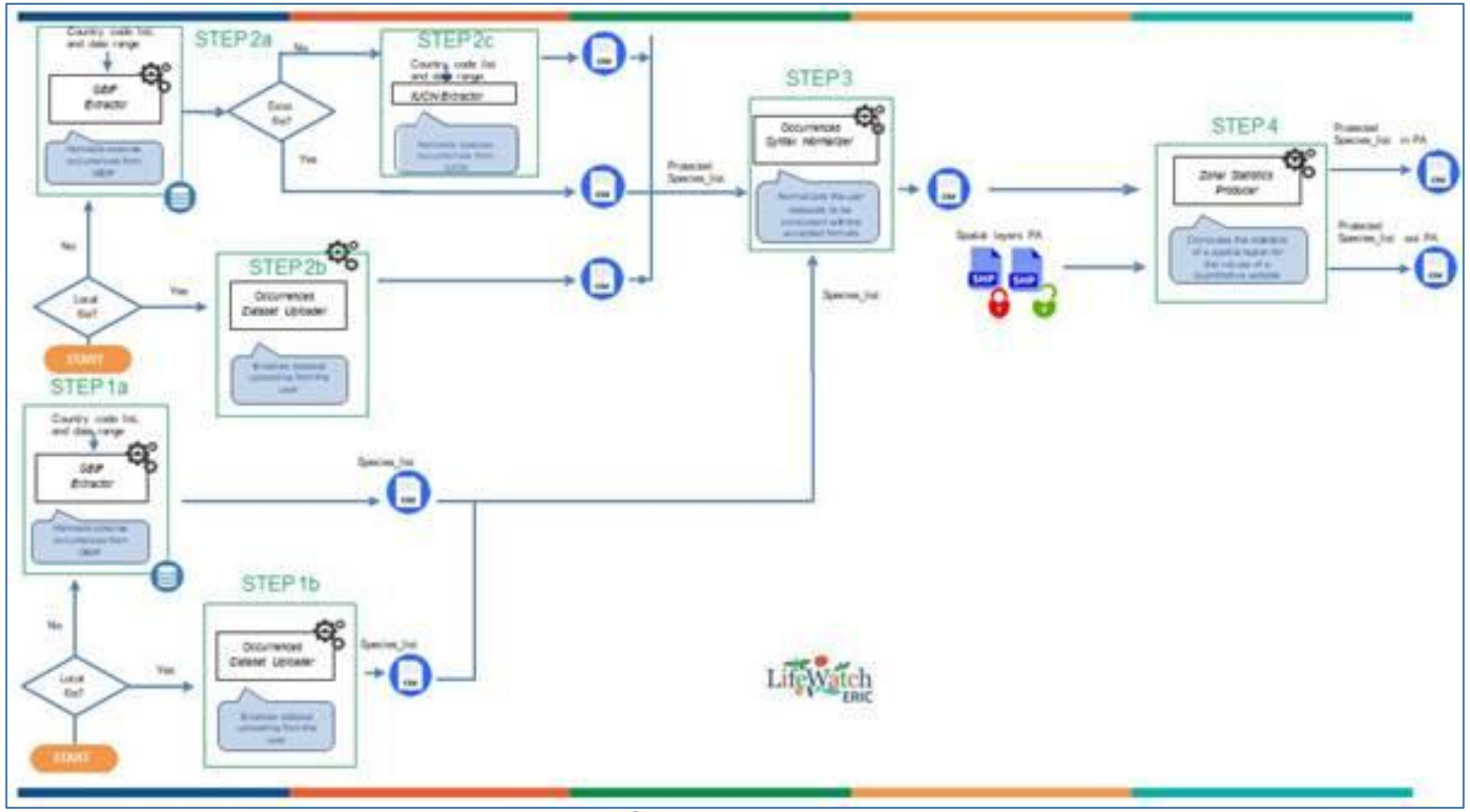
https://ipt.gbif.org/manual/en/ipt/latest/dwca-guide

https://www.gbif.org/es/tool/2yz6d3eCETWIF0eS7myRDf/frictionless-darwin-core



“Workflows” = flujo de trabajo

Automatización de proceso donde cada paso tiene sus propias reglas, y el resultado de un paso es la entrada para el siguiente



En diciembre de 2022, 196 naciones firmaron el histórico **Marco Mundial de Biodiversidad Kunming-Montréal (KM-GBF)**

El KM-GBF va acompañado de un **marco de seguimiento** (Monitoring Framework)que consta de un conjunto de **indicadores** para medir el progreso hacia los Objetivos y las Metas del Marco

Indicador (target 12): proporción de especies amenazadas que están adecuadamente conservadas dentro de las áreas protegidas

Par pasar a la discusión

- Todo esta conectado, y las conexiones van a más
    - Identificadores PIDs
    - Del Darwin Core Archive al Firctionless data
    - Frictionless data facilita la construcción de workflows
    - Darwin Core se basa en Dublin Core; Humboldt Core se basa en Darwin Core,…
- Construimos sistemas de contenidos, especificaciones (estándares) y herramientas buscando que cada vez nos sea más facil construir: para conocer más, para gestionar mejor

Francisco Pando

GBIF.ES, Unidad de Coordinación
Real Jardín Botánico - CSIC
Plaza de Murillo, 2
28014 Madrid, Spain

pando@gbif.es

Carlos de Mier. https://www.myxotropic.org/diachea-2/#gallery_fancybox_10192-1

MINISTERIO DE CIENCIA E INNOVACIÓN

CSIC

Gbif.es

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