

Tecnologías emergentes en el manejo de datos de biodiversidad

Francisco Pando
GBIF España



**Gobierno
de Canarias**



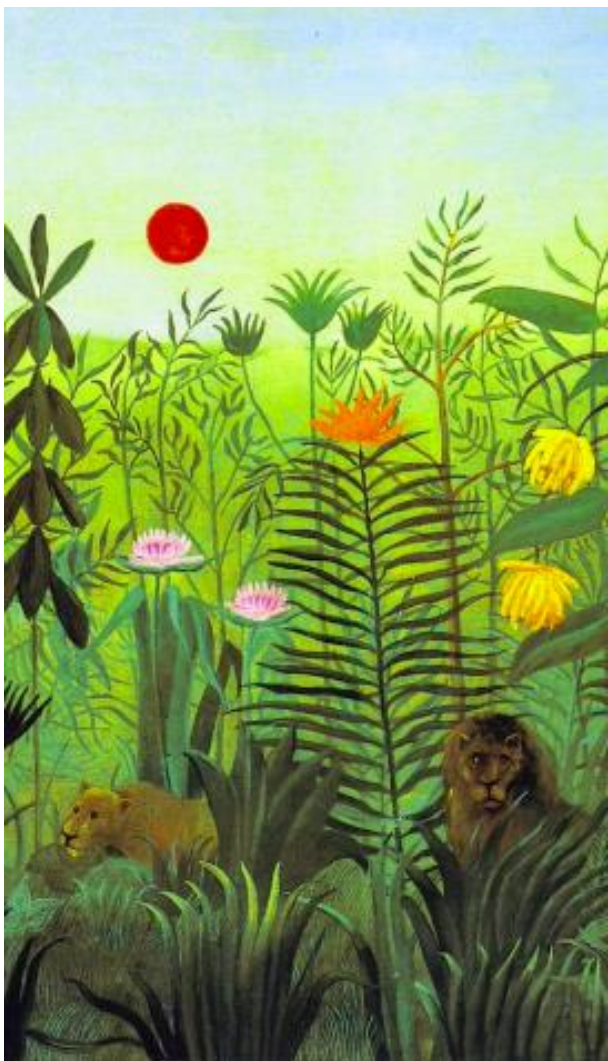
Banco de Datos de
Biodiversidad de
Canarias



EXCMO. AYUNTAMIENTO DE
SAN CRISTÓBAL DE
LA LAGUNA



Que vamos a ver



- Saber como se usan los datos
- Un poco de contexto, el TDWG
- Identificadores persistentes, otra vez
- Los DOIs, no solo para artículos científicos
- Como la IA va expandiéndose y especializándose
- Más allá de los principios FAIR: Interoperabilidad semántica
- Los datos viajan en paquetes
- Representando y conectando taxonomías: Checklist Bank
- Un estándar TDWG para seguimientos: Humboldt Core

Más allá de la biodiversidad y los datos

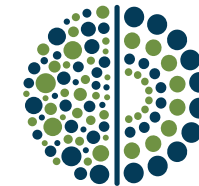
Science Review



<https://www.gbif.org/science-review>

AGRICULTURE
BIODIVERSITY SCIENCE
BIOGEOGRAPHY
CITIZEN SCIENCE
CLIMATE CHANGE
CONSERVATION
DATA MANAGEMENT
DNA
ECOLOGY
ECOSYSTEM SERVICES
EVOLUTION
FRESHWATER
HUMAN HEALTH
MARINE
PHYLOGENETICS
SPECIES DISTRIBUTION
TAXONOMY
INVASIVES

- La parte técnica (estándares, metadatos)

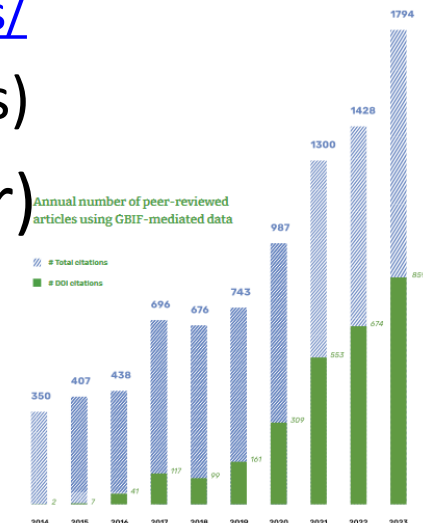


Biodiversity
Information
Standards

TDWG

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

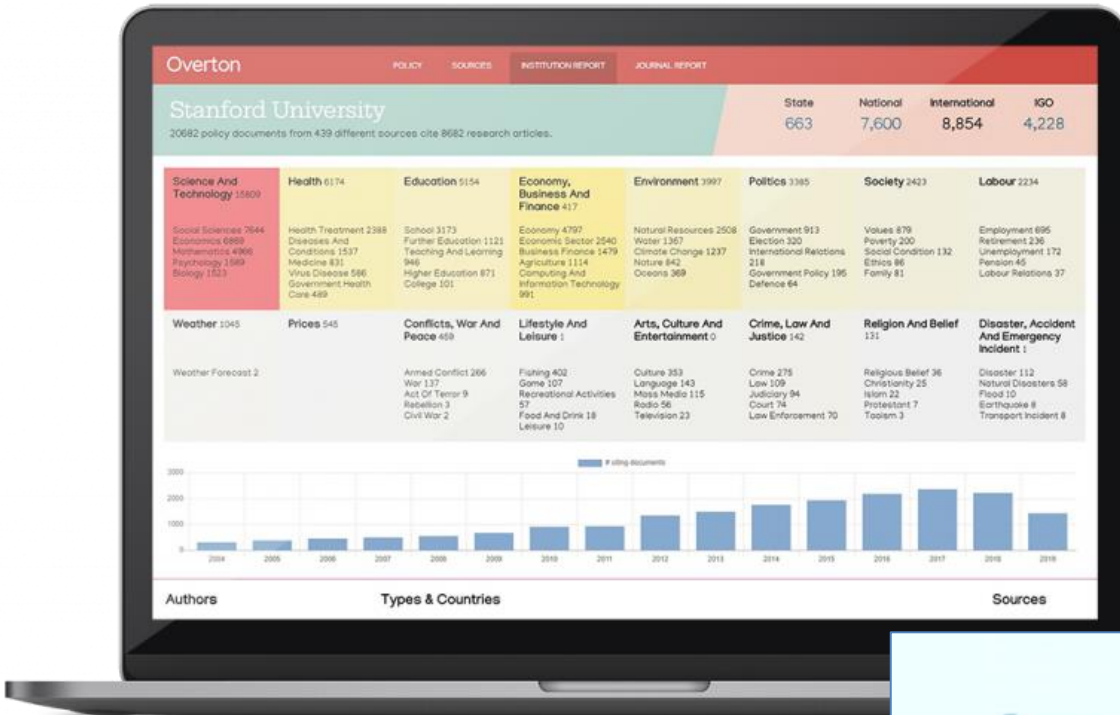
- Herramientas
 - Portal de publicadores
 - <https://ipt.gbif.es/>
 - Portal de usuarios)
- Cultura (usar, citar)



Overton

Overton es el índice de búsqueda más grande del mundo de documentos de políticas, directrices, publicaciones de grupos de expertos y documentos

<https://www.overton.io>



Overton

The world's largest searchable index of policy documents, guidelines, think tank publications and working papers

- ~1,400 GBIF-relevant documents identified, published by 350 bodies including more than 100 national, regional and municipal governments
- Other top contributors include IPBES, IUCN, Arctic Council, IOC-UNESCO and FAO



Government of Spain



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](#)

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

Biodiversity Information Standards (TDWG)
We are a non-profit organization and a community dedicated to developing biodiversity information standards.
65 followers | <https://www.tdwg.org> | @tdwg

Overview | Repositories 62 | Projects | Packages | People 23

Popular repositories

Repository Name	Description	Language	Stars	Forks	Visibility
dwc	Darwin Core standard for sharing of information about biological diversity.	Python	180	71	Public
dwc-qa	Public question and answer site for discussions about Darwin Core	Shell	46	8	Public
bdq	Biodiversity Data Quality (BDQ) Interest Group	HTML	43	6	Public
camtrap-dp	Camera Trap Data Package (Camtrap DP)	HTML	35	4	Public
wgsrpd	World Geographical Scheme for Recording Plant Distributions (WGSRPD)	HTML	34	28	Public
cd	Collection Descriptions	Python	24	10	Public archive

Repositories

<https://github.com/tdwg>

Identificadores persistentes: PIDs

- Únicos
- Universales
- Persistentes
- Resolubles



Despega con su uso para publicaciones

Diseñado para ser usado por máquinas

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.

GBIF lo adoptó para juegos de datos y descargas

Uso en expansion a otras áreas

- Resolver DOIs es gratis
- Acuñar DOIs cuesta

Identificadores persistentes: DOIs

“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”



Scientific advice for the determination of an EU-wide 2040 climate target and a greenhouse gas budget for 2030–2050

PUBLICATION INFORMATION
Published on 15 June 2023
Catalogue number: TH-03-23-229-EN-N
DOI: 10.2800/609405
ISBN: 978-92-9480-584-3



<https://data.europa.eu/doi/10.2800/609405>

Page 2/110

JUEGO DE DATOS DE REGISTROS BIOLÓGICOS | REGISTRADO

Localizaciones satélite de un ejemplar ("Lorien") de águila azor perdicera (*Aquila fasciata* Vieillot, 1822) en Aragón

Publicado por [Banco de Datos de la Biodiversidad en Aragón. Gobierno de Aragón](#)
Sanz Trullén V M

2896 REGISTROS 57 CITAS

CONJUNTO DE DATOS ESTADÍSTICAS ACTIVIDAD

Conjunto de posicionamientos de un ejemplar de águila azor perdicera (*Aquila fasciata* Vieillot, 1822), denominado "Lorien", equipado con un emisor ARGOS/GPS por el Servicio de Biodiversidad del Gobierno de Aragón. Los datos corresponden al periodo de actividad del emisor que abarca desde la instalación en agosto de 2011 hasta la muerte del ejemplar en agosto de 2012.

GOBIERNO DE ARAGON
Fecha de publicación: 30 de noviembre de 2018
Última modificación de metadatos: 19 de julio de 2021
Alojado por: GBIF-Spain
Licencia: CC BY 4.0

Cómo citar [DOI](#) 10.15470/joyvc1

57 citas

literatura, cómo se descubre y cómo se con los datos mediados de GBIF.

[DOI10.15470/joyvc1](https://doi.org/10.15470/joyvc1)

8876 eventos de descarga

Descargar reporte

JUEGO DE DATOS DE REGISTROS BIOLÓGICOS | REGISTRADO

Banco de Datos de la Biodiversidad de la Comunitat Valenciana

Publicado por [Biodiversity data bank of Generalitat Valenciana](#)
Conselleria de Medio Ambiente, Agua, Infraestructuras y Territorio. Generalitat Valenciana

2.619.112 REGISTROS 1459 CITAS

CONJUNTO DE DATOS PROYECTO ESTADÍSTICAS ACTIVIDAD

El Banco de Datos de Biodiversidad de la Comunidad Valenciana, es la mayor plataforma virtual de recopilación de datos actuales sobre la distribución geográfica de las especies silvestres de la Comunidad Valenciana, que valida, por expertos en cada uno de los grupos taxonómicos y hace públicas todas las citas que son recogidas por la ciudadanía, investigadores e investigadoras, universidades, administraciones públicas y otros organismos de interés, para poder completar el conocimiento del estado... Más

BDB
BANCO DE DATOS DE BIODIVERSIDAD
ID del proyecto: BDBCV

Fecha de publicación: 9 de mayo de 2024
Última modificación de metadatos: 9 de mayo de 2024
Alojado por: GBIF-Spain
Licencia: CC BY-NC 4.0

Cómo citar [DOI](#) 10.15468/b4yqdy

1459 citas

literatura, cómo se descubre y cómo se con los datos mediados de GBIF.

[DOI10.15468/dl.c8md9c](https://doi.org/10.15468/b4yqdy)

258.734 eventos de descarga

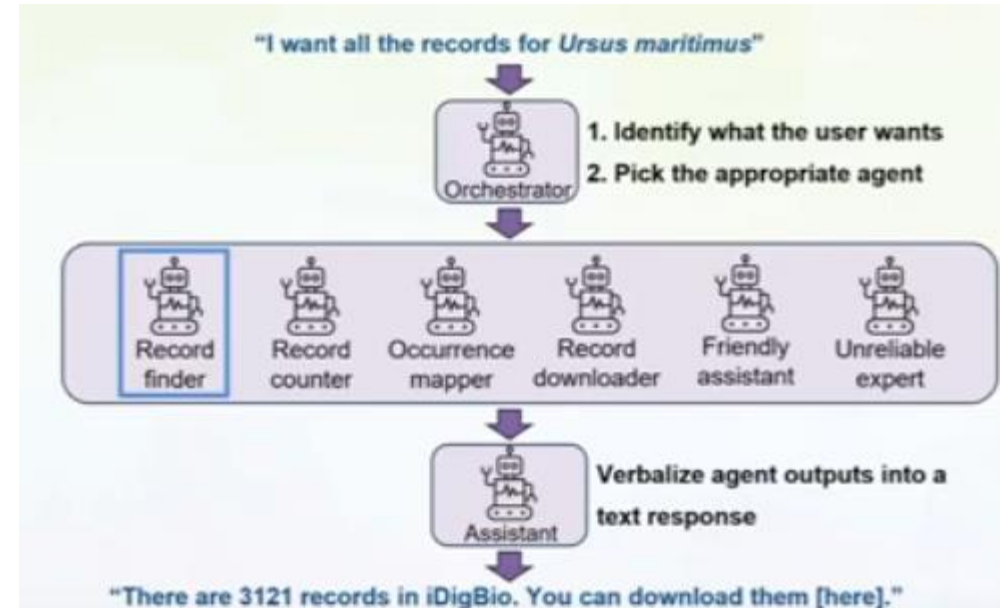
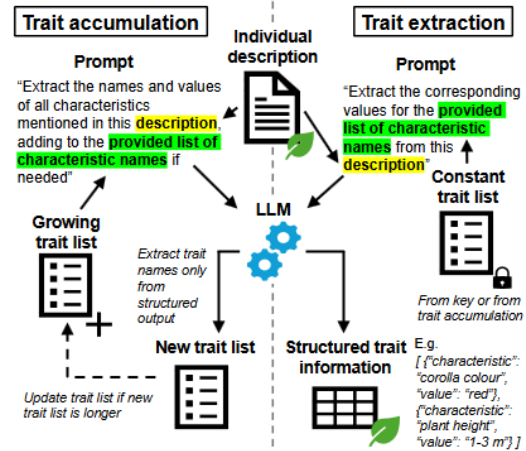
Descargar reporte

Inteligencia Artificial en la investigación en biodiversidad

Automated trait extraction from unstructured species descriptions for species ID: a pilot study using a large language model

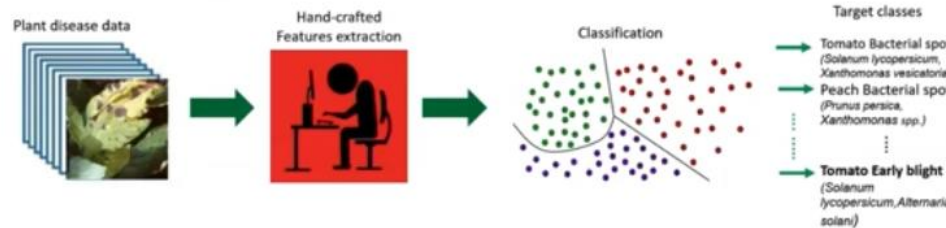
Y. J. Lee¹, M. Sosef², E. Lucas¹, V. Ung³, K. Gill¹, N. Nicolson¹

<https://zenodo.org/records/13691936>



Automated Plant Disease Recognition

Traditional Machine Learning



[Revolutionizing Plant Pathogen Conservation: The Past, Present, and Future of AI in Preserving Natural Ecosystems](https://virtual.oxfordabstracts.com/event/6771/session/104281)

<https://virtual.oxfordabstracts.com/event/6771/session/104281>

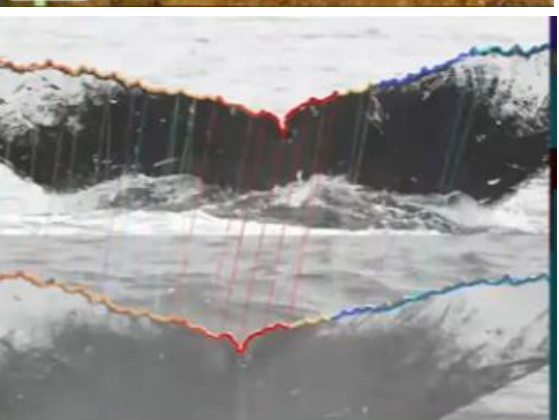
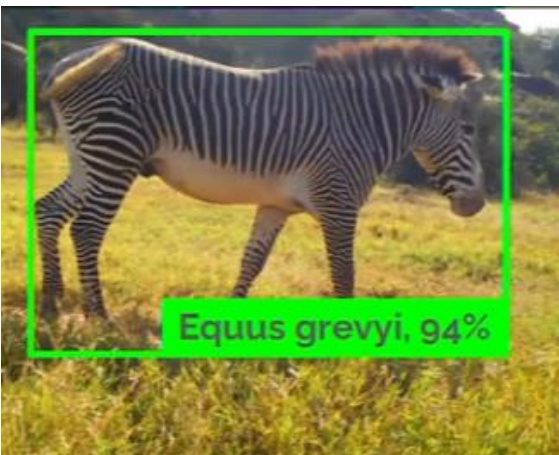
[Integrating LLMs and the iDigBio portal for conversational data exploration and inference](#)

- Method 1: Fully Open Source without Large Language Models (LLMs)
- Method 2: Fully Open Source with LLMs
- Method 3: Hybrid Approach (Open Source + API)
- Method 4: Fully Integrated LLMs (GPT-4 via API)
- Focus: Performance, cost, and time

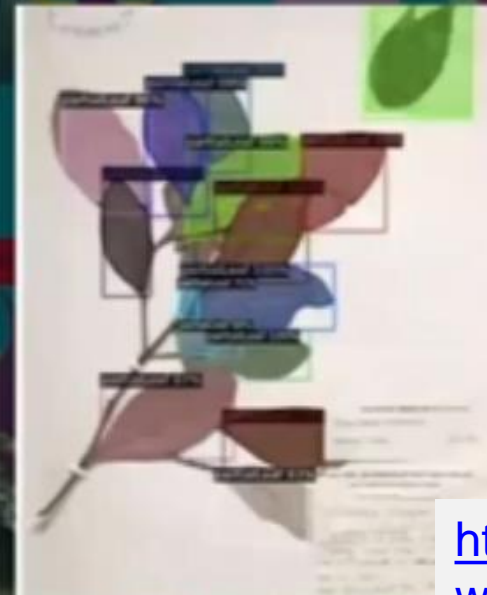
Kristen Lewers

[Leveraging LLMs to further understanding and engagement with Biodiversity Information Science and Standards through a customized fine-tuned chat model](#)

“Computer vision”



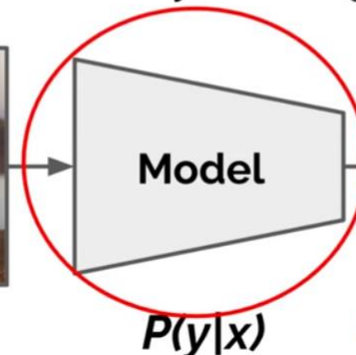
- Recognize species
- Identify individuals
- Count large groups
- Analyze behavior
- Monitor environment
- Measure traits



Sara Beery, MIT

<https://www.youtube.com/watch?v=QRGyJglF7Pk>

Which class y is in image x ?

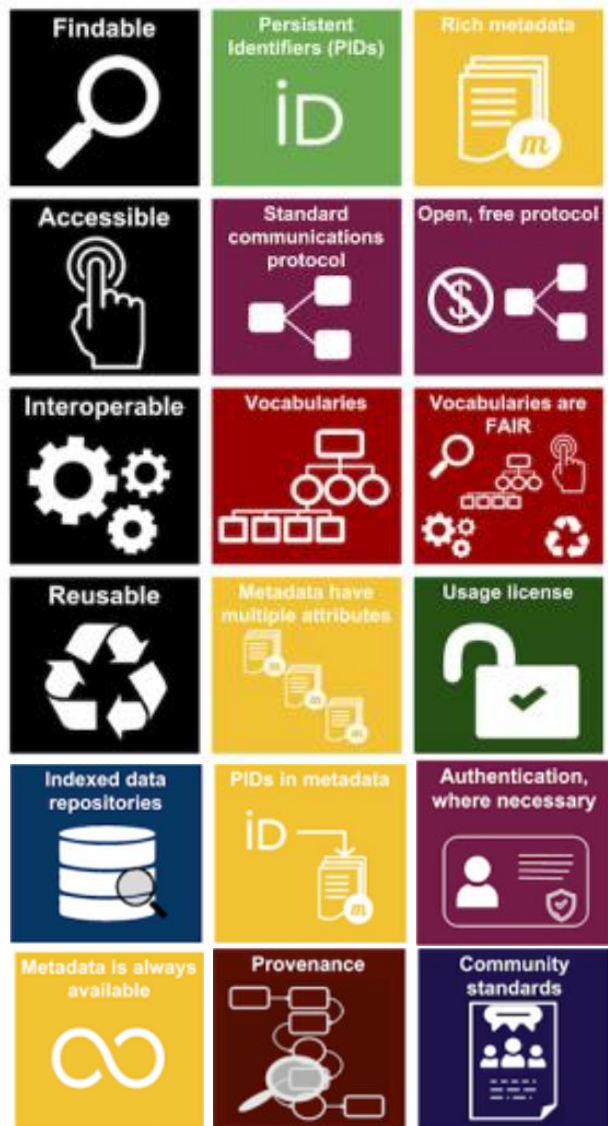


95% Elephant
5% Giraffe

Supervised CV
Models learn to
estimate $P(y|x)$ from
data

<https://cv4ecology.caltech.edu>

Principios FAIR e Interoperabilidad semántica



https://dwc.tdwg.org/list/#dwc_occurrenceID

Term Name dwc:occurrenceID	
Term IRI	http://rs.tdwg.org/dwc/terms/occurrenceID
Modified	2023-06-28
Term version IRI	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28
Label	Occurrence ID
Definition	An identifier for the dwc:Occurrence (as opposed to a particular digital record of the dwc:Occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the dwc:occurrenceID globally unique.
Notes	Recommended best practice is to use a persistent, globally unique identifier.
Examples	http://arctos.database.museum/gus:MSB:Mamm:23362-000866d2-c177-4648-a200-ead4007051b9

Metadata about this term version are available in the following formats/serializations:

Description	IRI
HTML file (this document)	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28.htm
RDF/Turtle	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28.ttl
RDF/XML	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28.rdf
JSON-LD	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28.json

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

Principios FAIR e Interoperabilidad semántica



<https://www.ands.gov.ar/with-data/fairdata>

Term Name	
Term IRI	
Modified	
Term version	
IRI	
Label	
Definition	
Notes	
Examples	



<https://www.ands.gov.ar/with-data/fairdata>

Estándares de datos: ahora y en el futuro

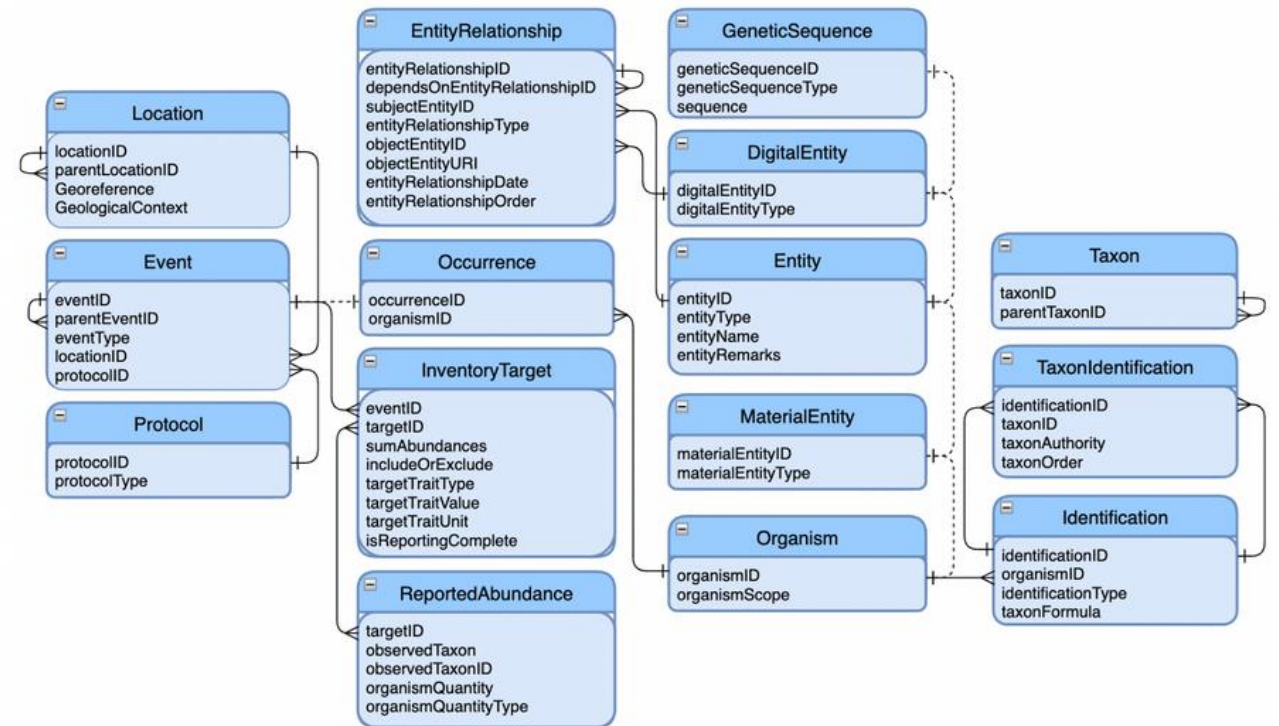
Especificación de datos:

Darwin Core (lista de propiedades)

Record-level
Occurrence
Organism
MaterialEntity
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship
UseWithIRI
LivingSpecimen
PreservedSpecimen
FossilSpecimen
MaterialCitation
HumanObservation
MachineObservation
Cite Darwin Core

“New data Model”

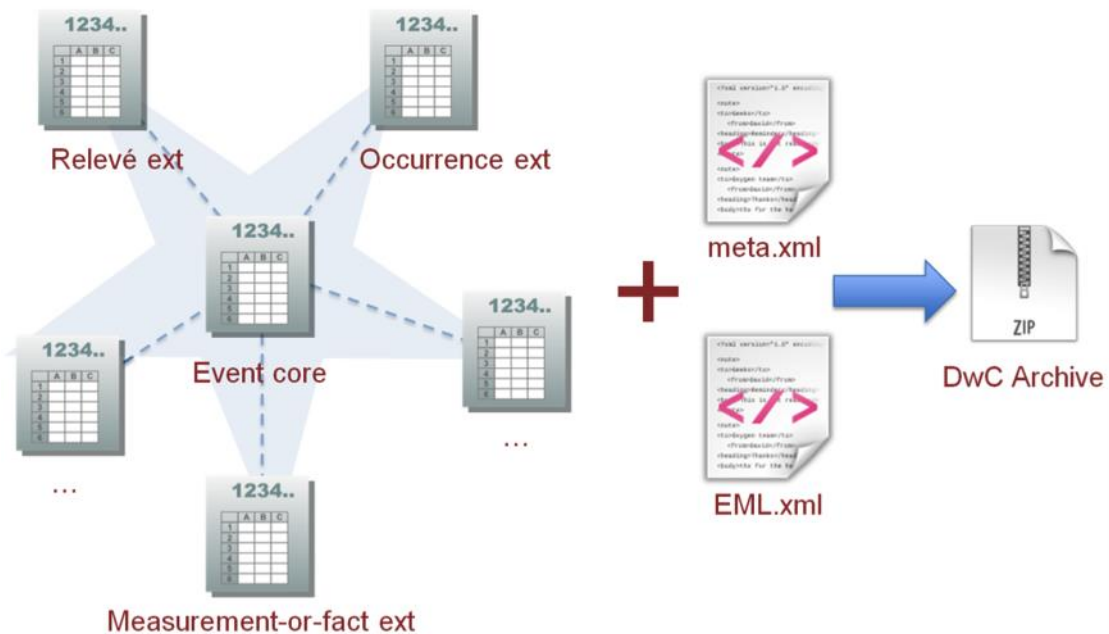
Conceptual model in a database



ER Diagram of the Unified Model with Auxiliary Tables (2024-03)

Formatos de intercambio: ahora y en el futuro

Darwin Core Archive



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

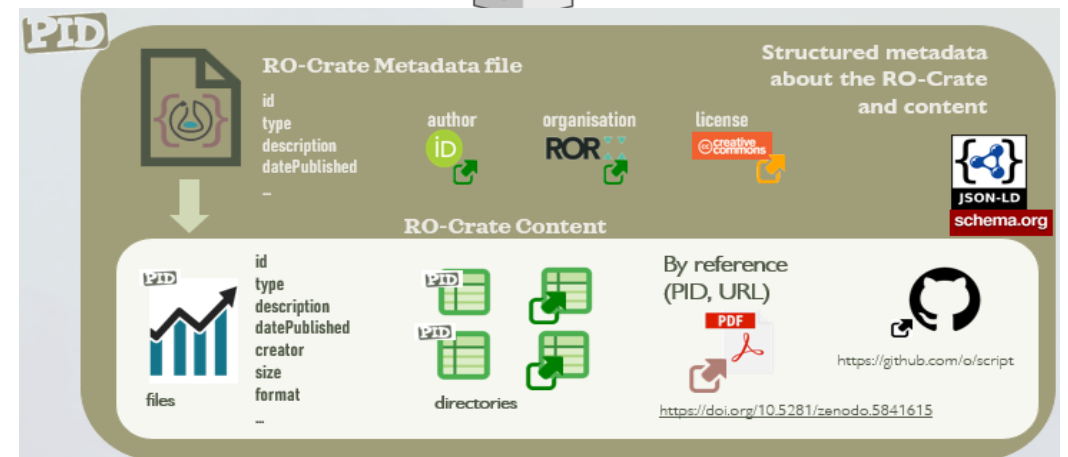
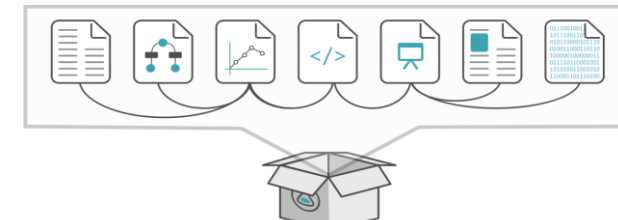
- Frictionless data Packages

<https://frictionlessdata.io/>

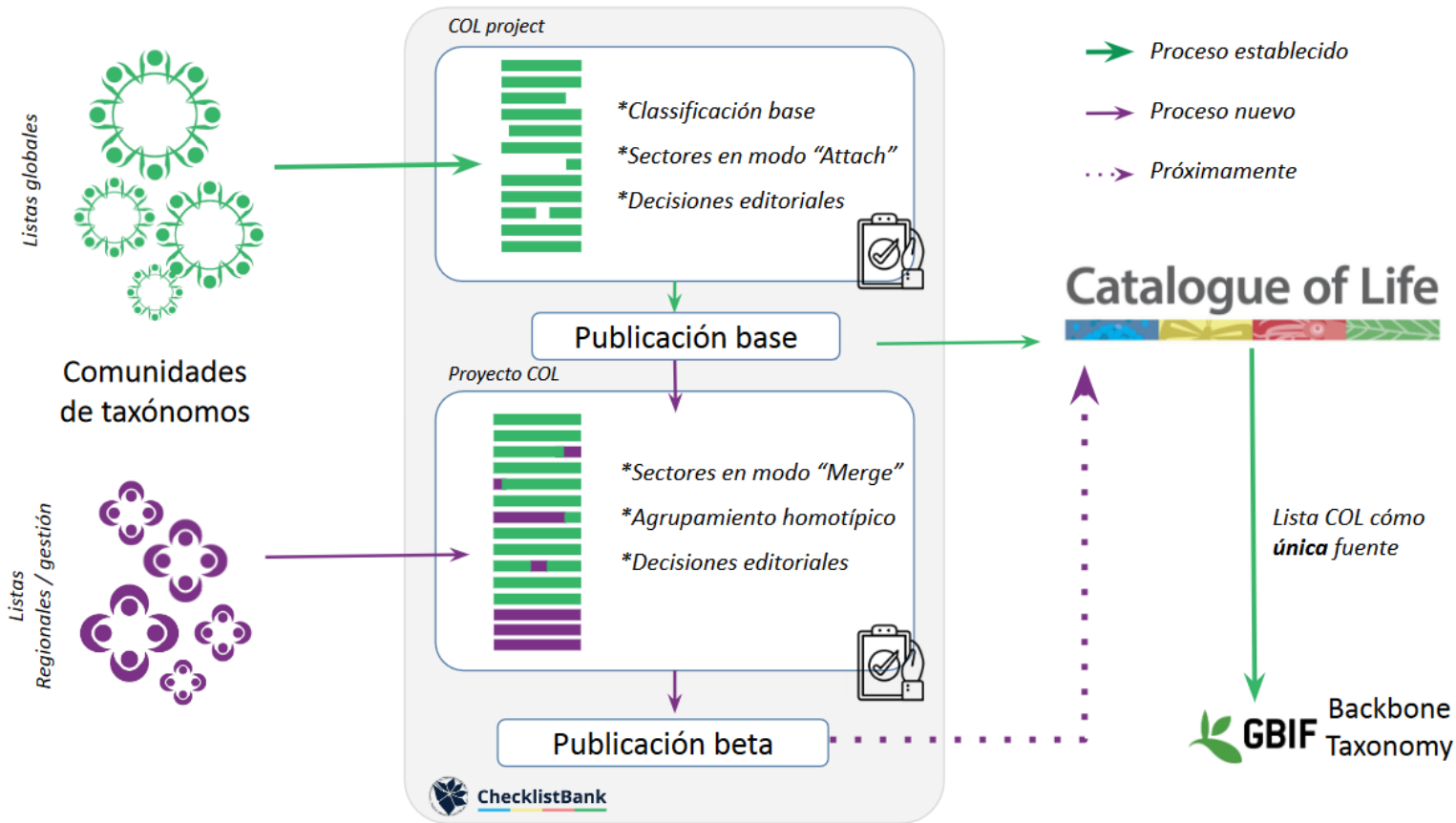
- RO-Crates

<https://www.researchobject.org/ro-crate/specification.html>

Describe and **package** data collections, datasets, software etc. **with their metadata Platform-independent** object exchange between repositories and services Support **reproducibility** and **analysis**: link data with codes and workflows



CheckList Bank



Alias	Title	Version	Logo	Creator	Edit
+ COL	Catalogue of Life	working draft		Bánki O., Rosko...	
COLH	Catalogue of Lif...	1.4		Ruggiero M. A....	
CCW	Catalogue of Cr...	May 2021	CCW	Oosterbroek P.	
CIPA	Computer Aide...	3, Mar 2011		Vignes-Lebbe R...	
ReptileDB	The Reptile Dat...	2024-03		Uetz P., Hošek J.	

Dataset	Scientific Name	Status	Rank	Parents
+ iNaturalist Taxonomy (2024-11-01)	<i>Quercus ilex</i>	accepted	species aggregate	Cerris > sect. Ilex
+ The World Checklist of Vascular Plants (WCVP) (13.0)	<i>Quercus ilex</i> L.	accepted	species	Fagaceae > Quercus
+ Catalogue of Life Checklist	<i>Quercus ilex</i> Lour.	ambiguous synonym of <i>Quercus helferiana</i> A.D.C.	species	Quercus > Quercus helferiana
+ Catalogue of Life Checklist (Annual Checklist 2018)	<i>Quercus ilex</i> Lour., nom. illeg.	synonym of <i>Quercus helferiana</i> A.D.C. sensu Govaerts	species	Quercus > Quercus helferiana
+ Flora Helvetica - Fagaceae	<i>Quercus ilex</i> L.	accepted	species	Fagaceae > Quercus

- Búsquedas en múltiples listas
- Comparar nombres entre listas
- Métricas de Calidad detalladas
- Facilitar la construcción de listas

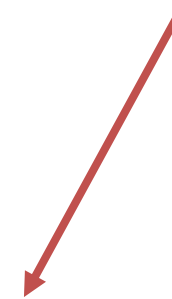
• Y contribuir

Humboldt Extension for Ecological Inventories

The Humboldt Extension for Ecological Inventories is a vocabulary for transmitting information about biodiversity surveys with hierarchical structure. It is used along with Darwin Core terms to extend descriptions of Events.

<https://eco.tdwg.org/>

Site
Habitat Scope
Temporal Scope
Taxonomic Scope
Organismal Scope
Methodology Description
Material Collected
Sampling Effort
UseWithIRI



A JOURNAL OF SPACE AND TIME IN ECOLOGY

Editorial | [Open Access](#)

Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment

Robert Guralnick ✉, Ramona Walls, Walter Jetz

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

Methodology Description

- compilationTypes
- compilationSourceTypes
- inventoryTypes
- protocolNames
- protocolDescriptions
- protocolReferences
- isAbundanceReported
- isAbundanceCapReported
- abundanceCap
- isVegetationCoverReported
- isLeastSpecificTargetCategoryQuantityInclusive

compilationTypes		Property
Identifier	http://rs.tdwg.org/eco/terms/compilationTypes	
Definition	A statement specifying whether data reported are derived from sampling events, ancillary data compiled from other sources, or a combination of both.	
Comments	This term is only relevant if the dwc:Event is an inventory. Recommended best practice is to use a controlled vocabulary. Recommended best practice is to separate the values in a list with space vertical bar space (). This term has an equivalent in the dwciri: namespace that allows only an IRI as a value, whereas this term allows for any string literal value.	

- Identificadores persistentes (DOIs, IRIs)

- Estándares

- IA, procesamiento de los datos en general, automatización

Trazabilidad, transparencia, impacto, economía

- Cuestiones científicas

- Retos sociales

- Gestión, conservación

Francisco Pando

GBIF.ES, Unidad de Coordinación
CSIC

Joaquín Costa, 22
28002 Madrid, Spain

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 Español 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>

