Jornadas sobre Información de Biodiversidad y Administraciones Ambientales

20 y 21 de noviembre de 2024, Antíguo Convento de Santo Domingo - La Laguna, Tenerife, Canarias

Tecnologías emergentes en el manejo de datos de biodiversidad

Francisco Pando
GBIF España

Gobierno de Canarias

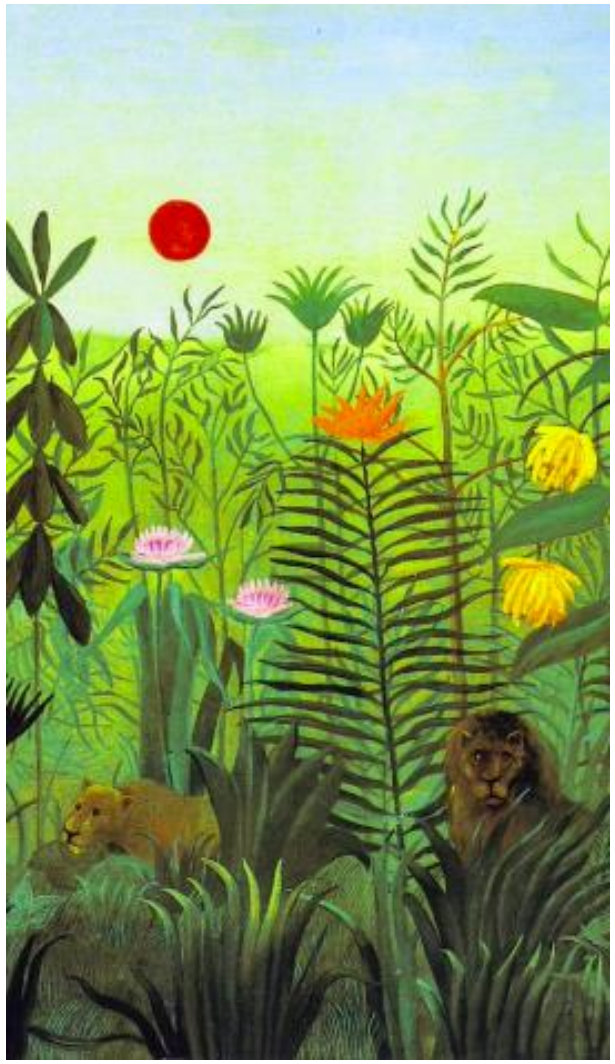
biota Banco de Datos de Biodiversidad de Canarias

EXCMO. AYUNTAMIENTO DE SAN CRISTÓBAL DE LA LAGUNA

GBIF.ES

CSIC

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

“Paisaje exótico con león y leona en Africa”. Henri Rousseau. https://wikioo.org .

Más allá de la biodiversidad y los datos

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

- La parte técnica (estándares, metadatos)

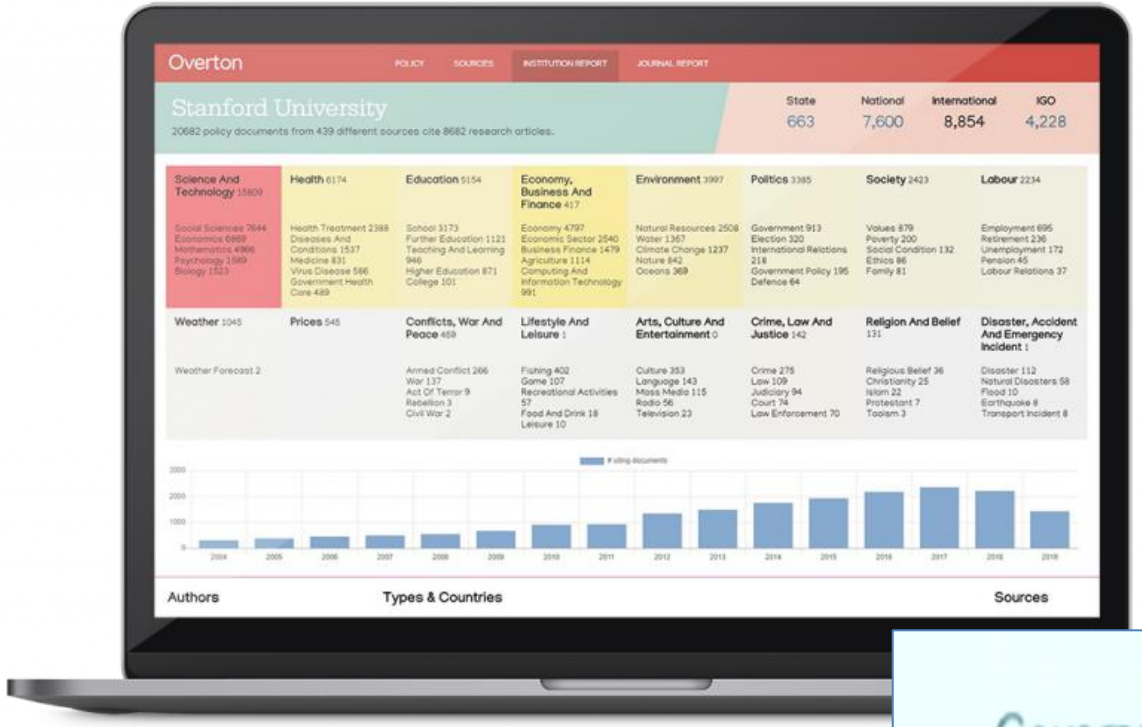
Biodiversity Information Standards TDWG https://www.tdwg.org/

- 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

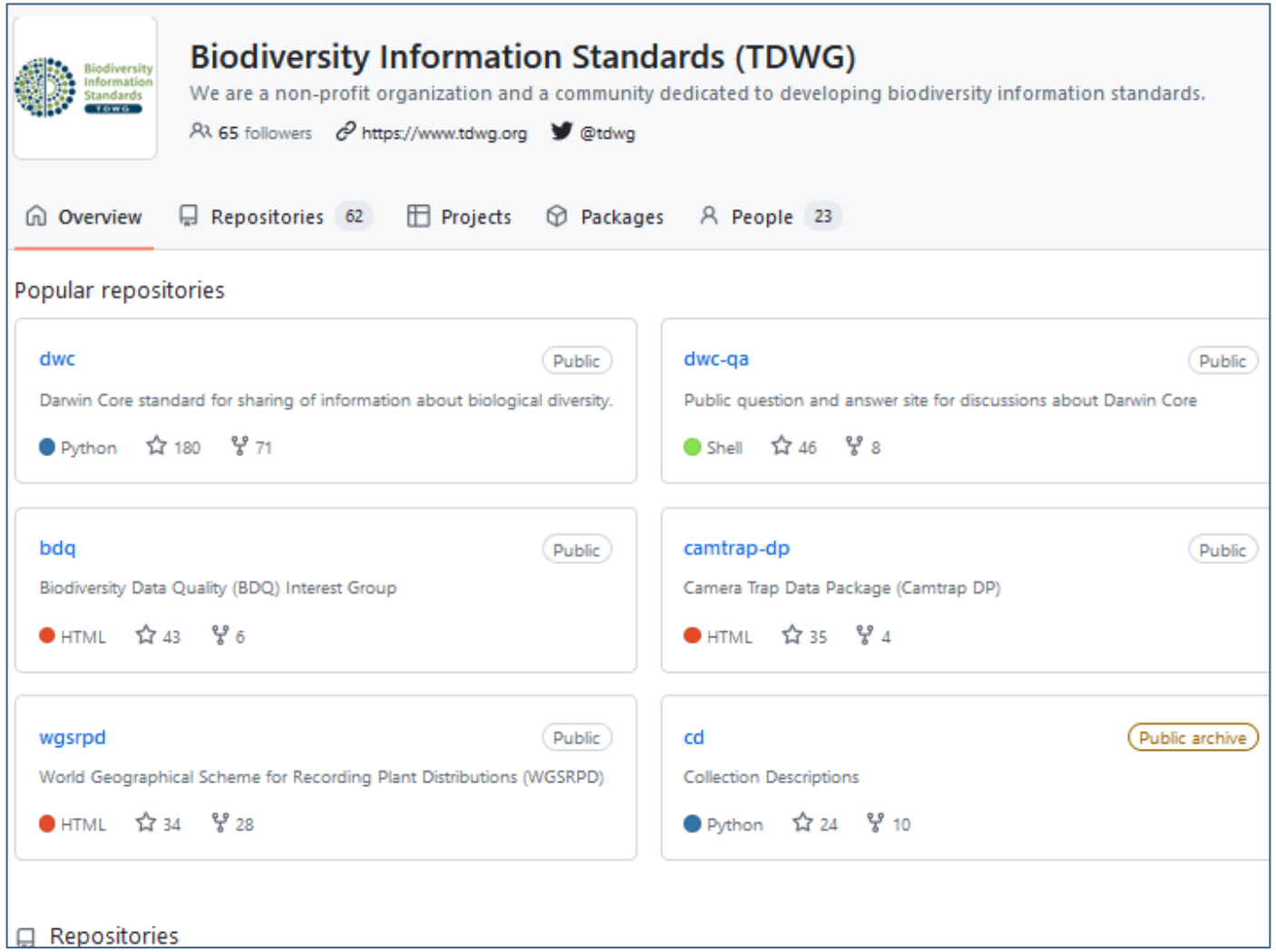
https://www.overton.io



TDWG: Biodiversity Information Standards

Biodiversity Information Standards TDWG

- Access to Biological Collection Data (ABCD) Schema
- Audiovisual Core Multimedia Resources Metadata Schen
- 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/

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.

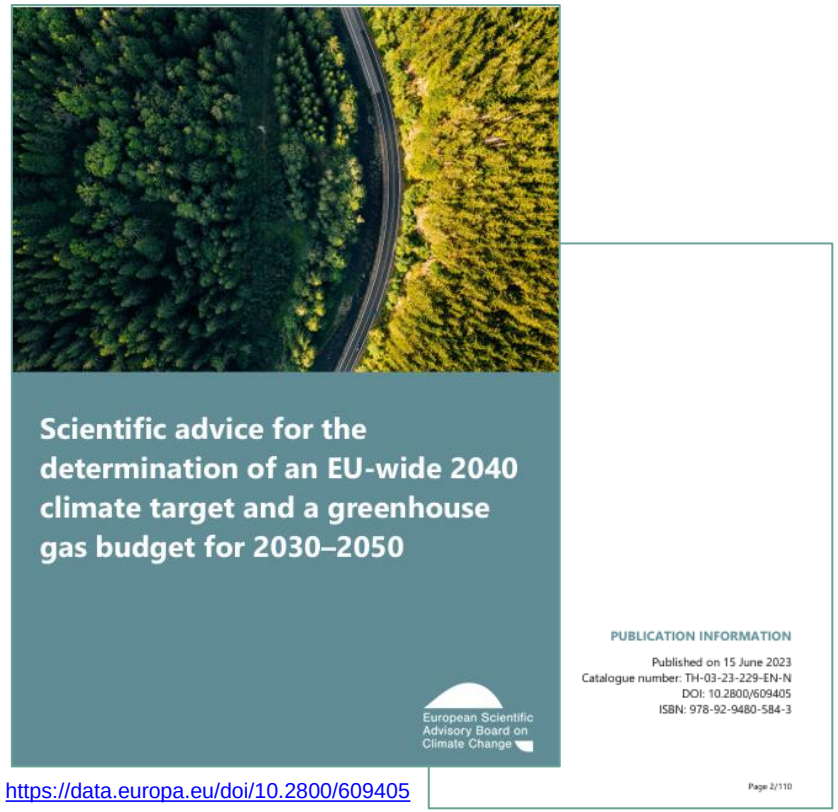
GBIF lo adoptó para juegos de datos y descargas

Uso en expansion a otras áreas

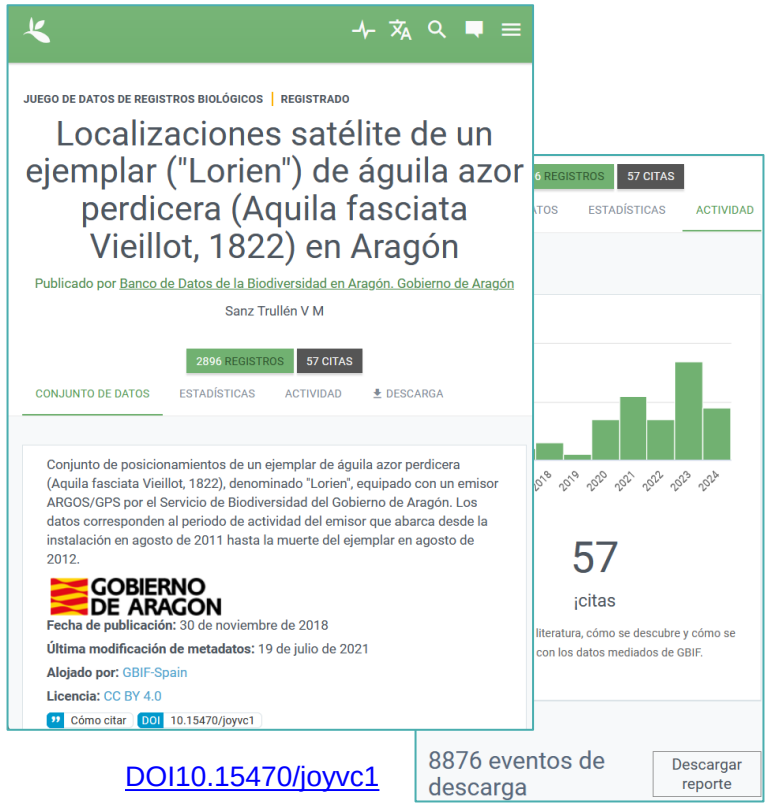
- Resolver DOIs es gratis
- Acuñar DOIs cuesta

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"



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



DOI10.15470/joyvc1

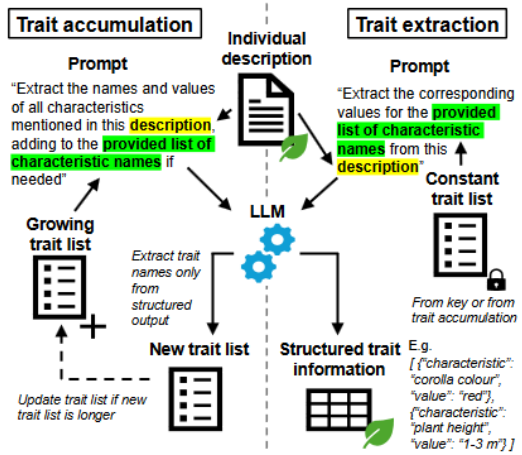
DOI10.15468/dl.c8md9c

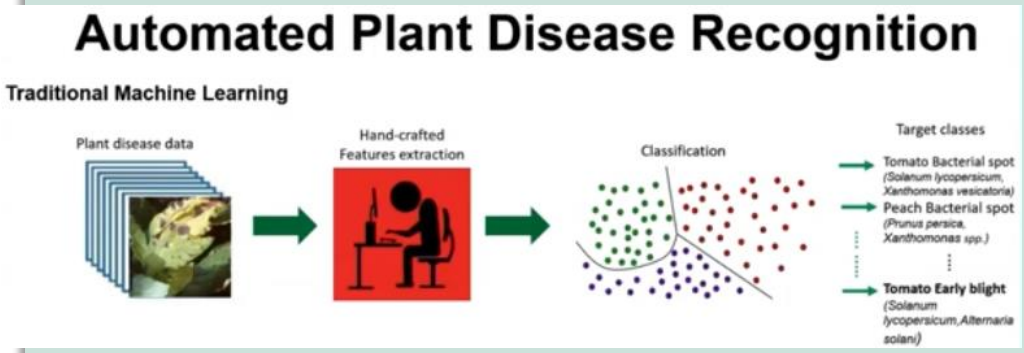
Inteligencia Artificial en la investigación en biodiversidad

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

Y. J. Lee [1], M. Sosef [2], E. Lucas [1], V. Ung [3], K. Gill [1], N. Nicolson [1]

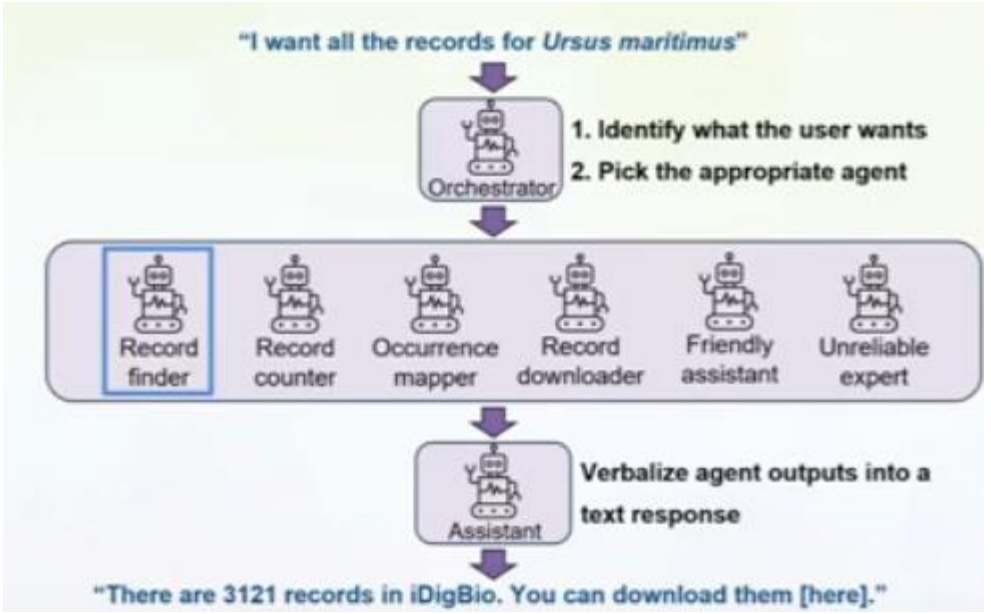
https://zenodo.org/records/13691936



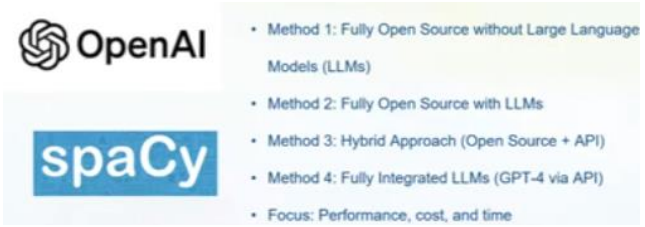


Revolutionizing Plant Pathogen Conservation: The Past, Present, and Future of AI in Preserving Natural Ecosystems

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



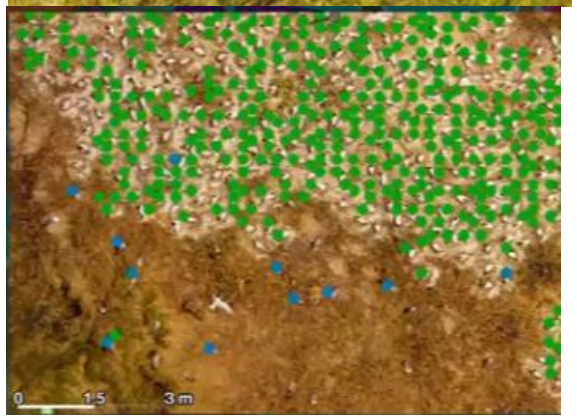
Integrating LLMs and the iDigBio portal for conversational data exploration and inference

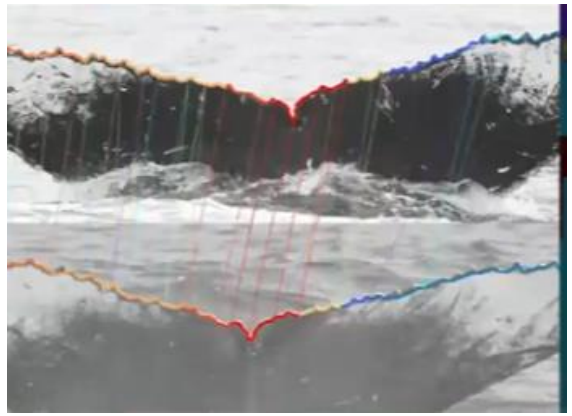


Kristen Lewers

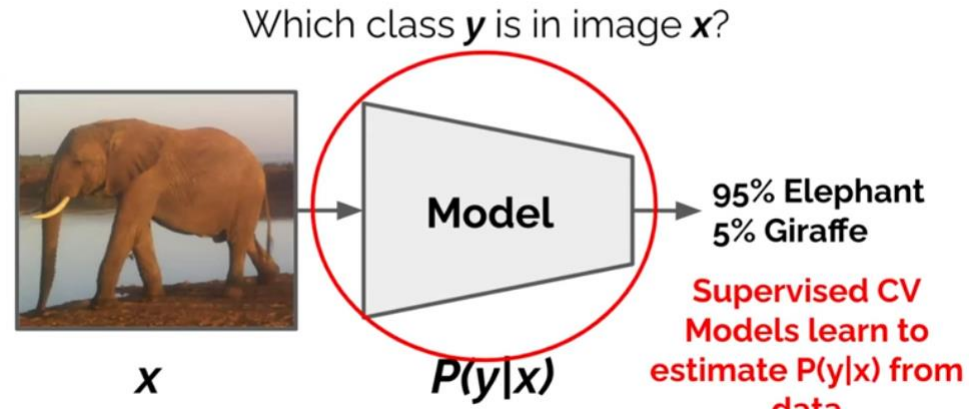
Leveraging LLMs to further understanding and engagement with Biodiversity Information Science and Standards through a customized fine-tuned chat model

“Computer visión”





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



https://cv4ecology.caltech.edu

Principios FAIR e Interoperabilidad semántica

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

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/guid/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

Principios FAIR e Interoperabilidad semántica

https://

Term Nam
Term IRI
Modified
Term versi
IRI
Label
Definition
Notes
Examples

https://www.ands
with-data/fairdata

Estándares de datos: ahora y en el futuro

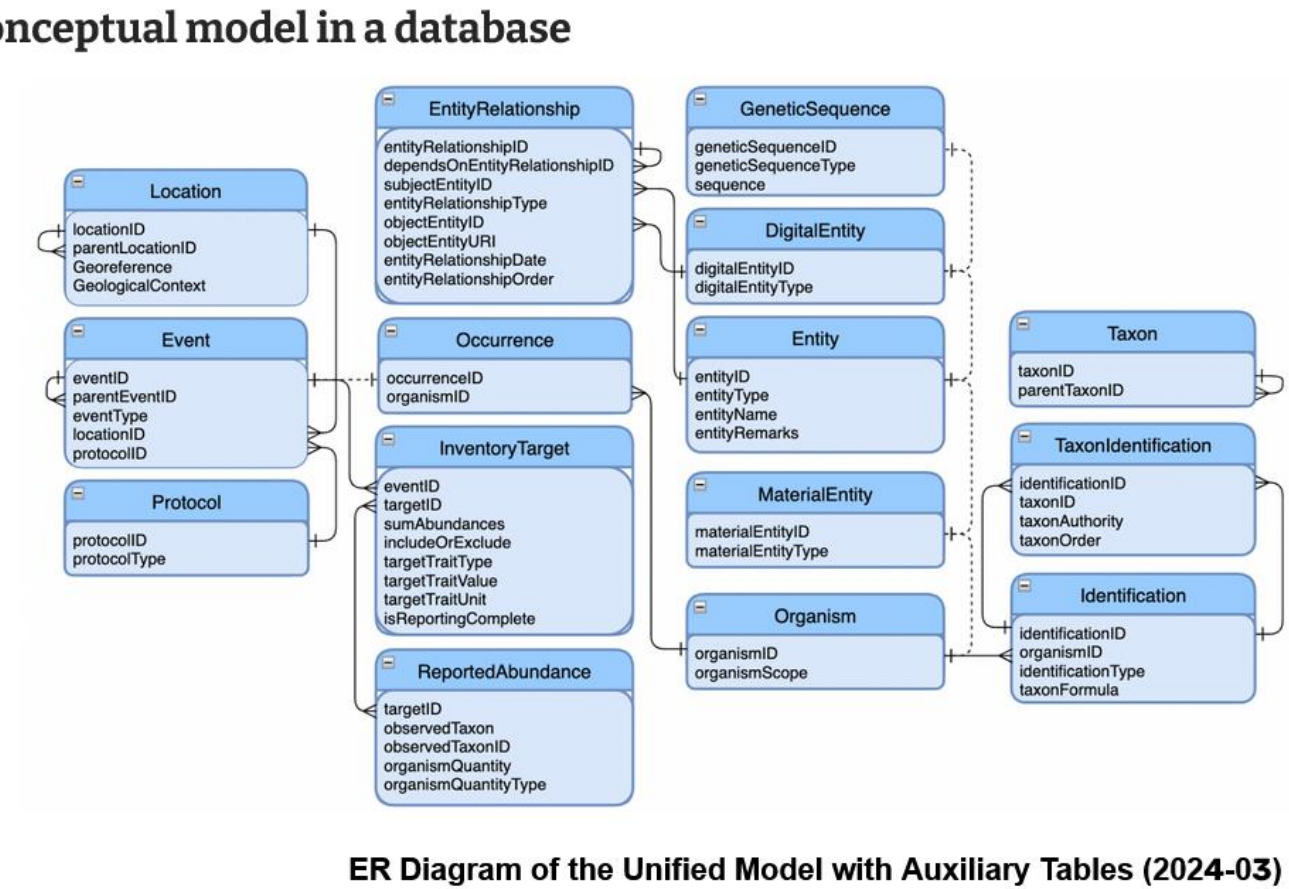
Especificación de datos:

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

https://dwc.tdwg.org/terms/

"New data Model"

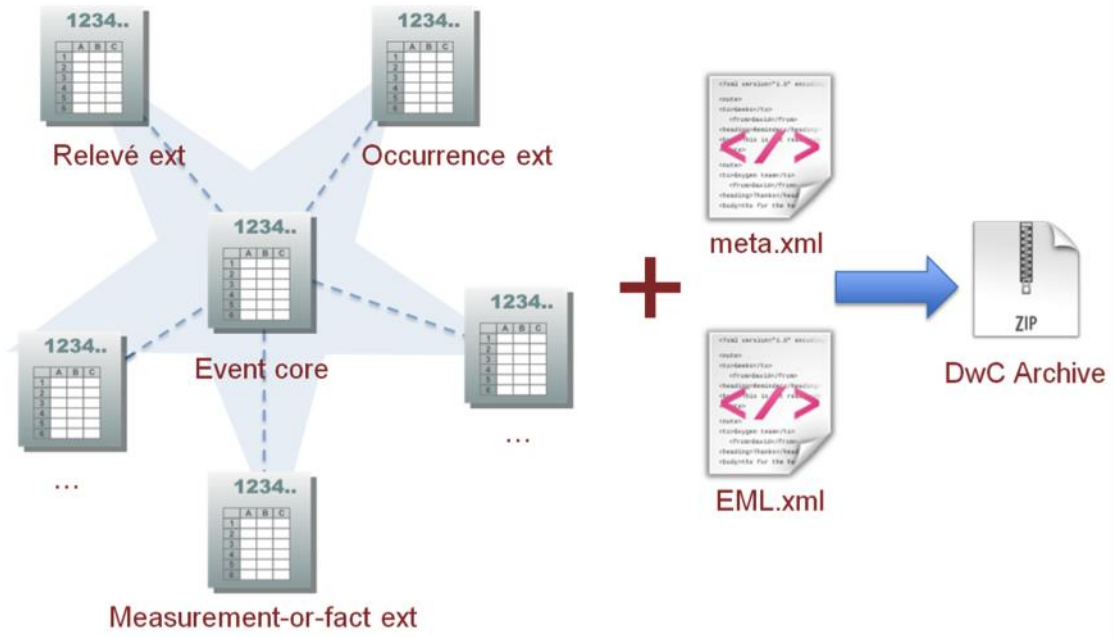


ER Diagram of the Unified Model with Auxiliary Tables (2024-03)

https://www.gbif.org/new-data-model

Formatos de intercambio: ahora y en el futuro

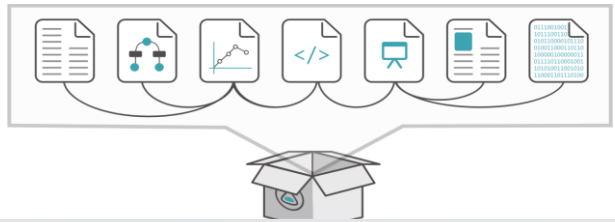
Darwin Core Archive

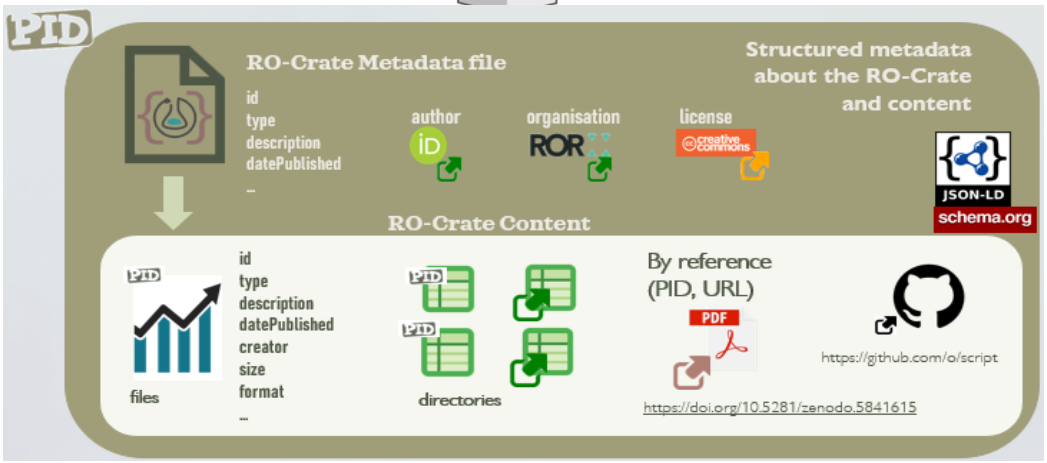


https://ipt.gbif.org/manual/en/ipt/latest/dwca-guide

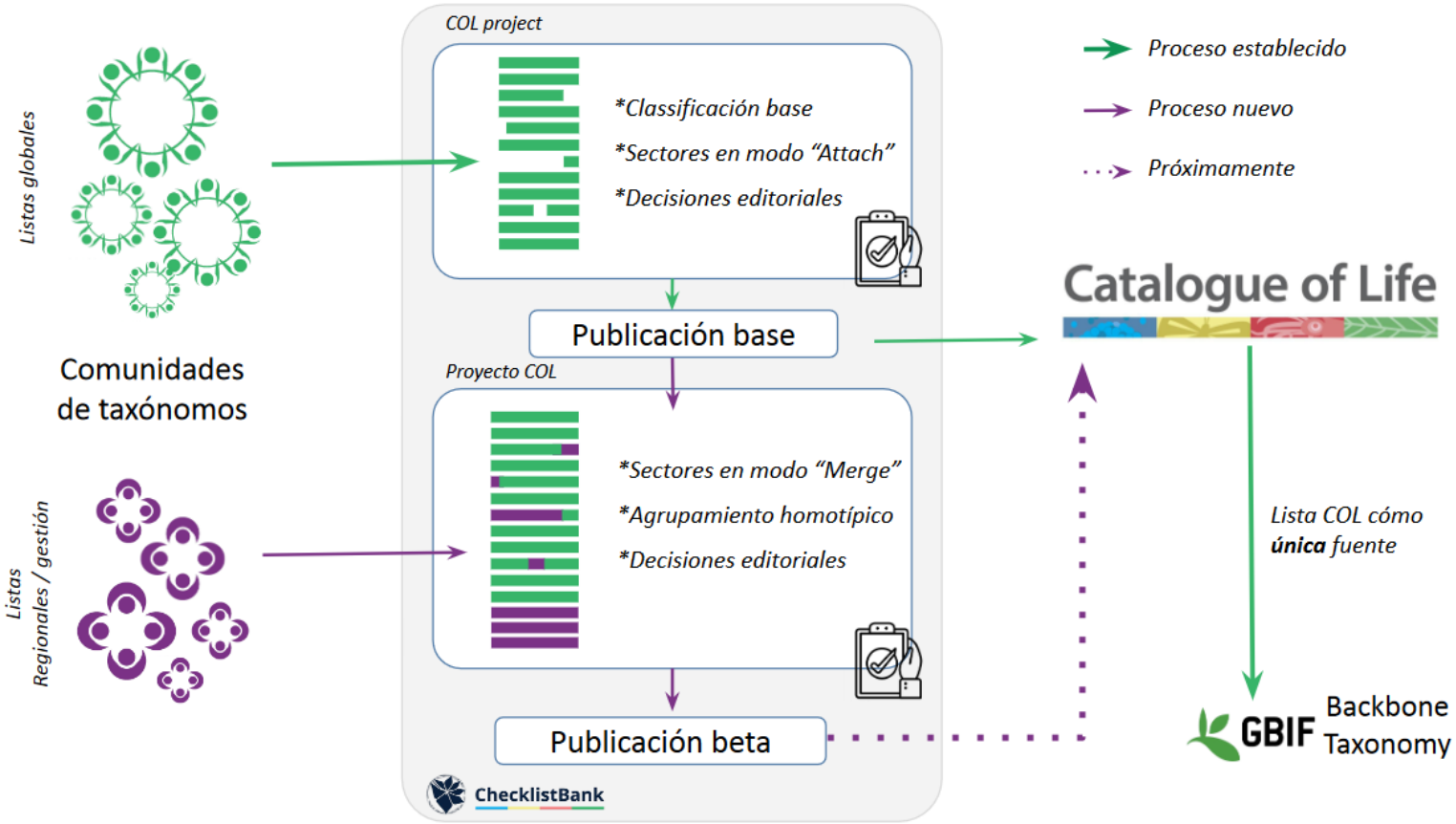
- Frictionless data Packages
  https://frictionlessdata.io/
- RO-Crates
  https://www.researchobject.org/ro-crate/specification.html

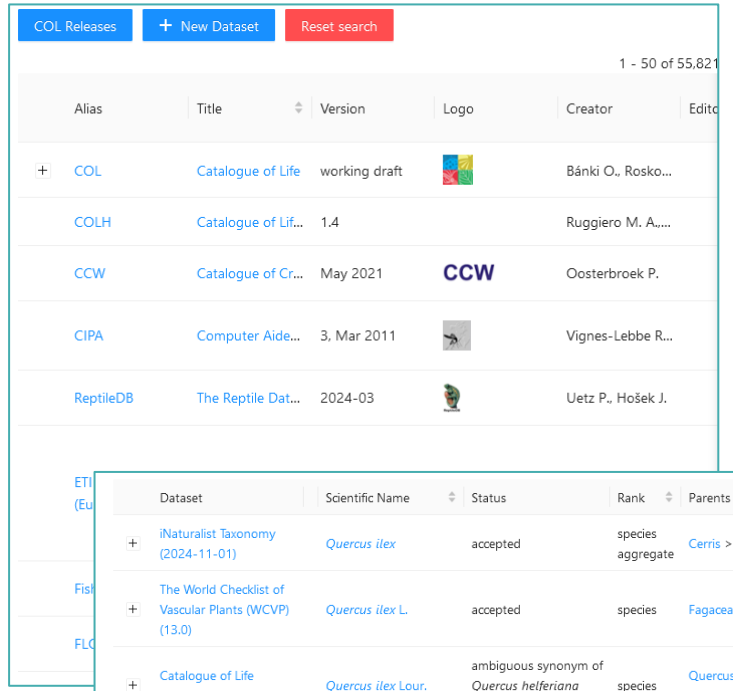
Describe and **package** data collections, datasets, software etc. **with their metadata Platform-independent** object exchange between repositories and services Support **reproducibility** and **analysis**: link data with codes and workflows

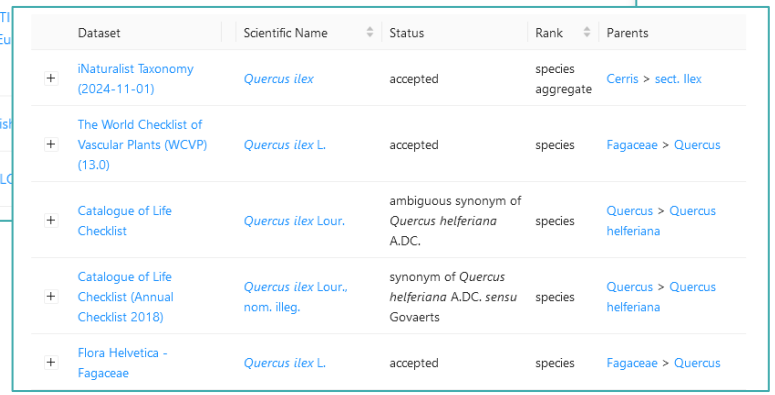




CheckList Bank







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

- Y contribuir

Diana Hernández-Robles & Camila Andrea Plata

https://gbif.es/talleres/webinar-gbif-es-plataformas-de-informacion-en-biodiversidad/

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

ECOGRAPHY Open Access
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
280o2 Madrid, Spain

pando@gbif.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

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

GBIF.ES