Bloque 4 - Documentación, Informatización y publicación

13. Tecnología y estrategias emergentes

Fancisco Pando -pando@gbif.es

III Taller GBIF.ES: Conservación, gestión e informatización de herbarios
Madrid - 10 – 12 de marzo de 2026
Katia Cezón, GBIF España

AHIM

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

CSIC

GBIF.ES

REAL JARDÍN BOTÁNICO

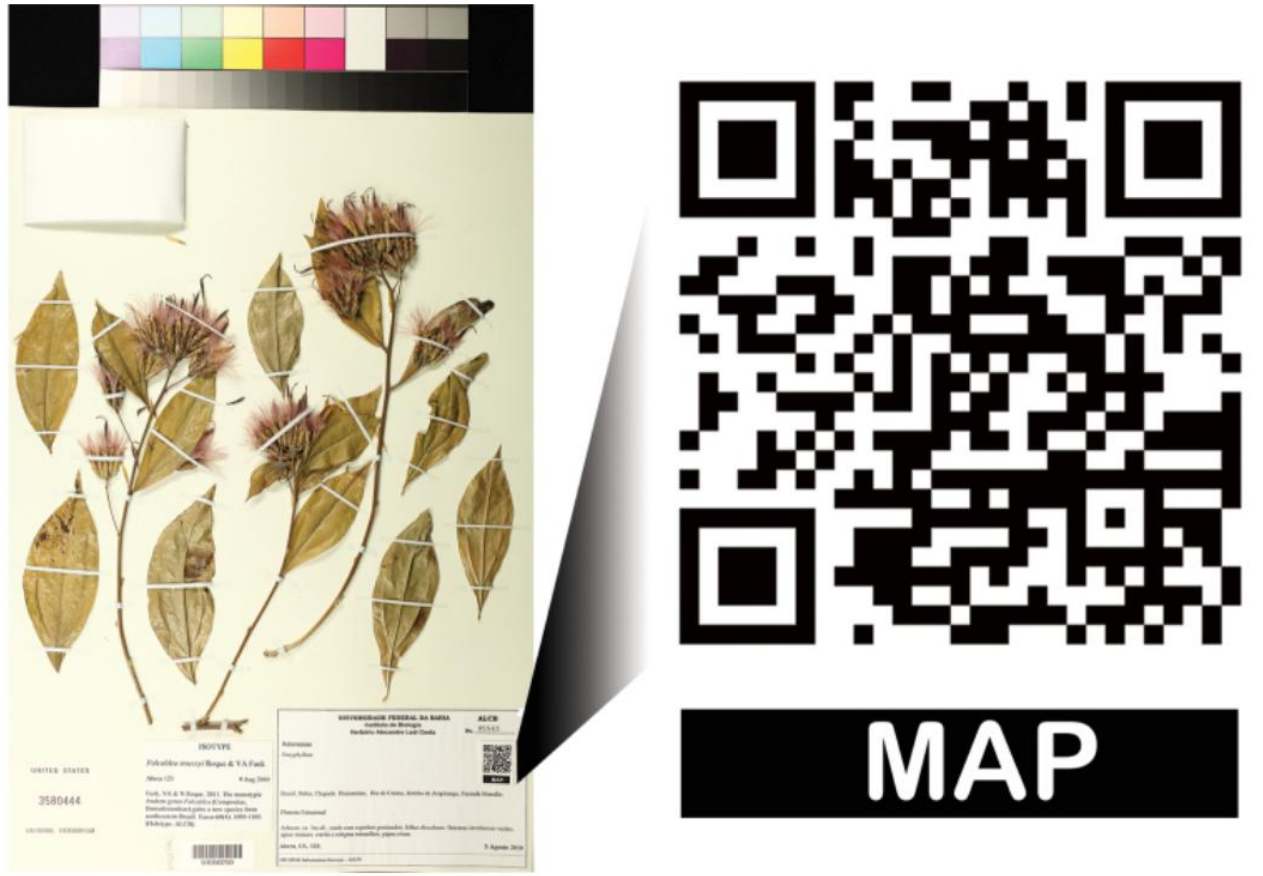
Que vamos a ver

- GBIF Unified model
- Darwin-DP
- Entornos virtuales de Investigación y colecciones
- Herbarios virtuales (sí, la AHIM tiene uno)
- Ciencia ciudadana y colecciones
- IA para etiquetas
- IAs

No nuevo: Códigos de barras vs QRs

Herbarium specimen of large flowered trillium (Trillium grandiflorum) with QR code on label that directs researchers to the associated iNaturalist observation, including images of the plant in the wild.

https://carnegiemnh.org/expanding-the-value-of-herbarium-specimens-with-citizen-science-app-inaturalist/



Example of plant specimen at the US herbarium with a QR code linking to a Google map of the collection locality. Encoded text: "http://bit.ly/130tnzO".

https://phytokeys.pensoft.net/articles.php?id=1478

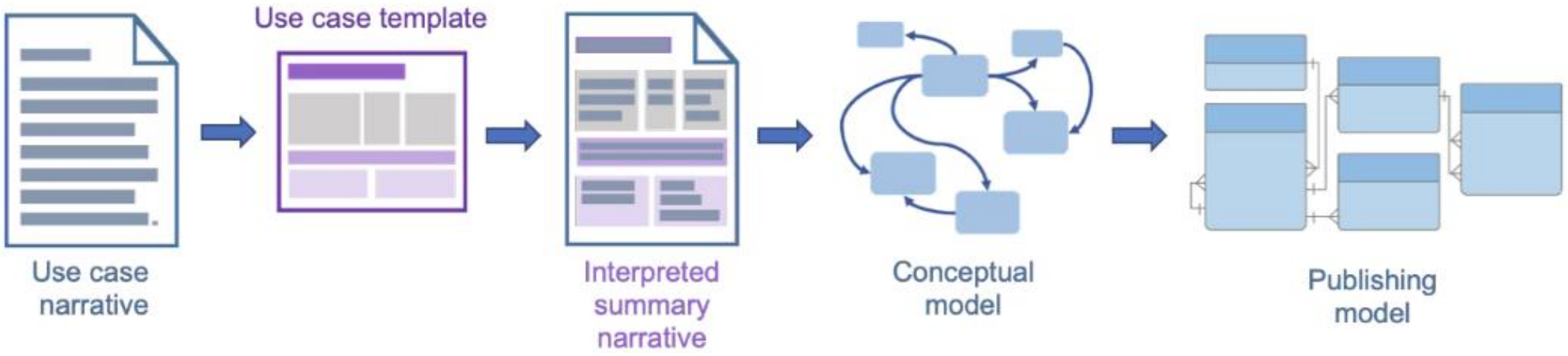
Diazgranados M, Funk V (2013) Utility of QR codes in biological collections. PhytoKeys 25: 21-34

Nuevo modelo conceptual para Darwin Core

>> define la estructura lógica y las relaciones entre los distintos elementos del estándar Darwin Core (DwC).Sobre el conjunto de términos o vocabulario controlado que es Darwin Core, el modelo conceptual describe cómo se conectan entre sí esos términos —es decir, la arquitectura conceptual que sustenta el intercambio de datos de biodiversidad.

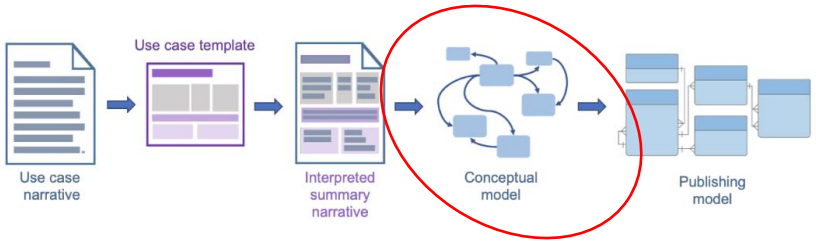
"...biodiversity data is more complicated than 'just' the occurrence of species in time and space; organisms interact, co-occur, move and evolve.
This implies a need for richer and more varied types of data than GBIF has thus far supported."

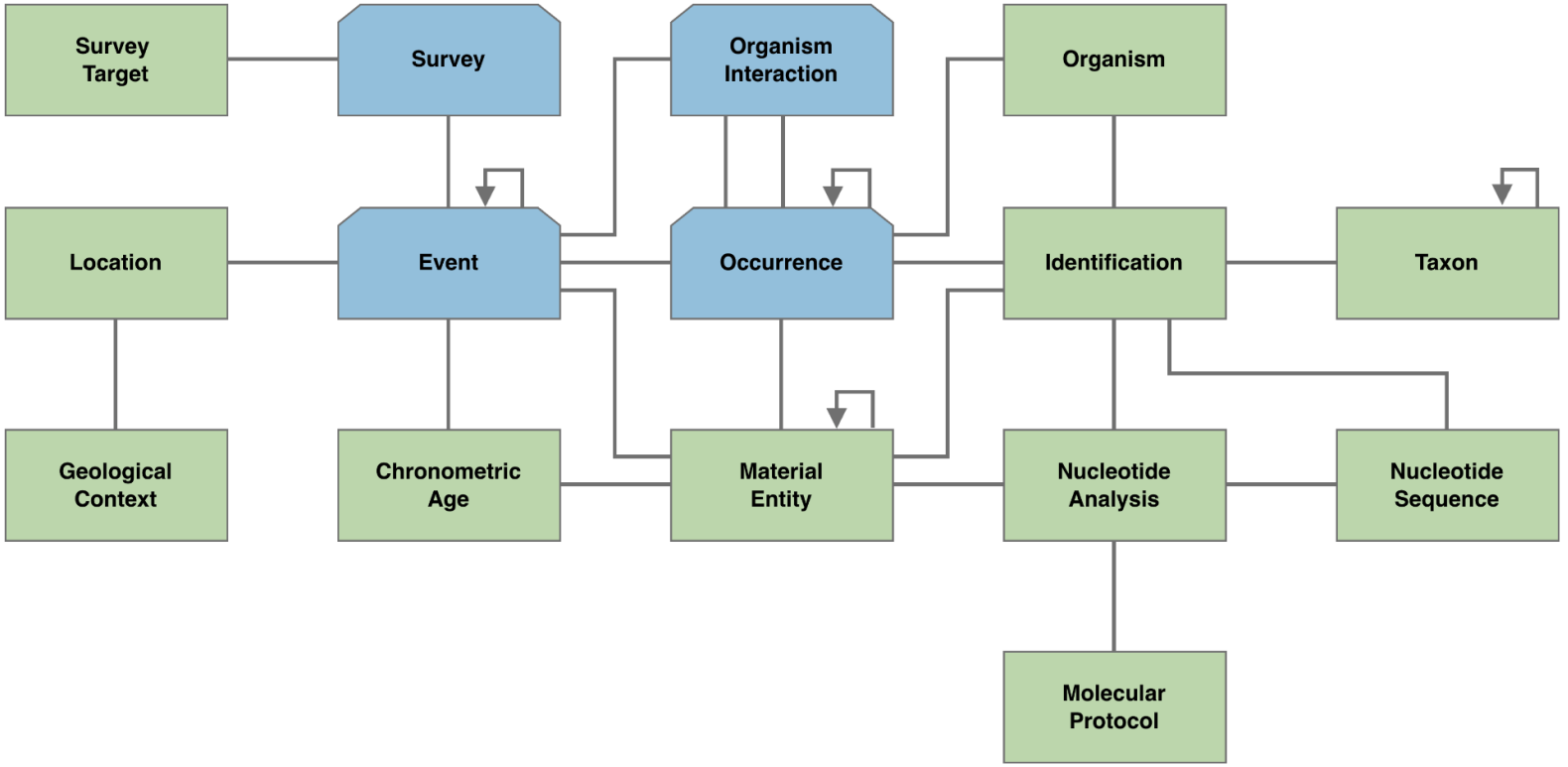
Heberling et al. (2021) Data integration enables global biodiversity synthesis
https://doi.org/10.1073/pnas.2018093118

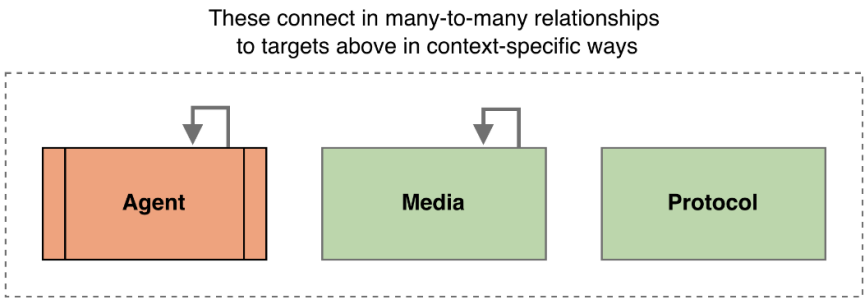


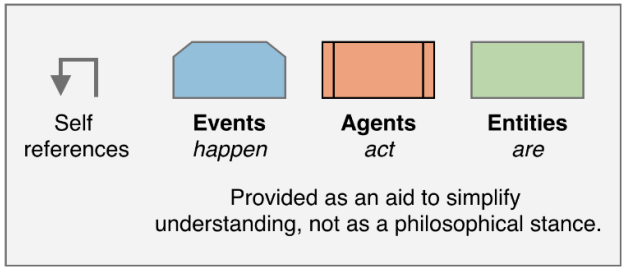
Nuevo modelo conceptual para Darwin Core

https://gbif.github.io/dwc-dp/cm/

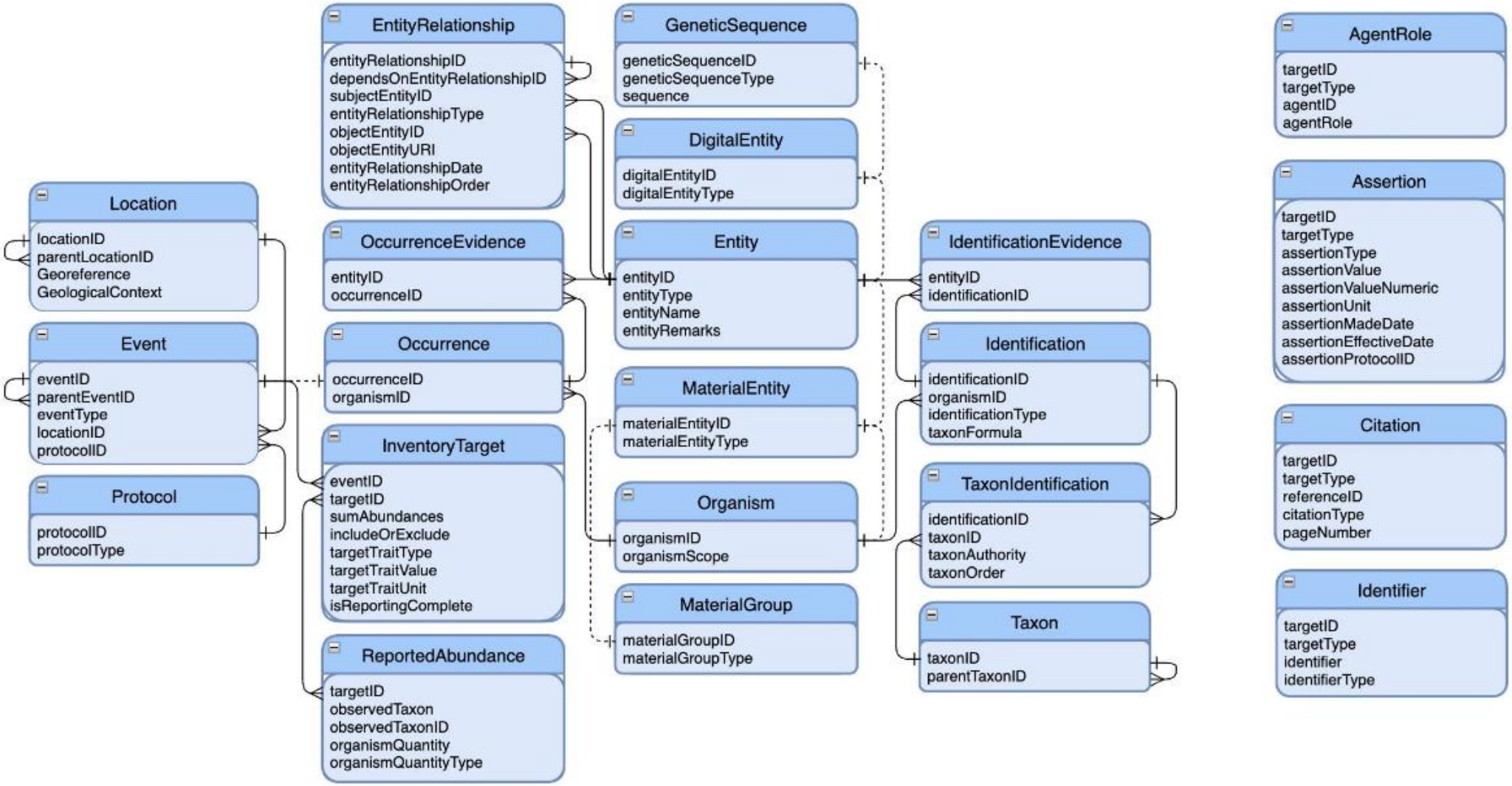




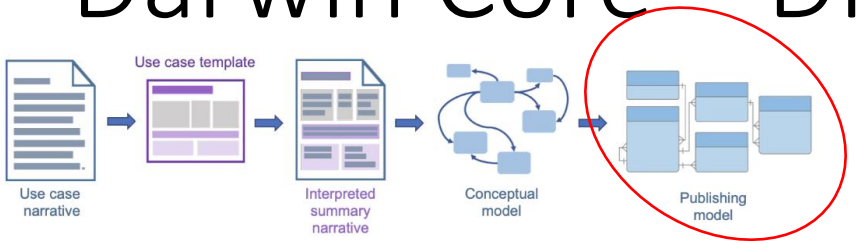


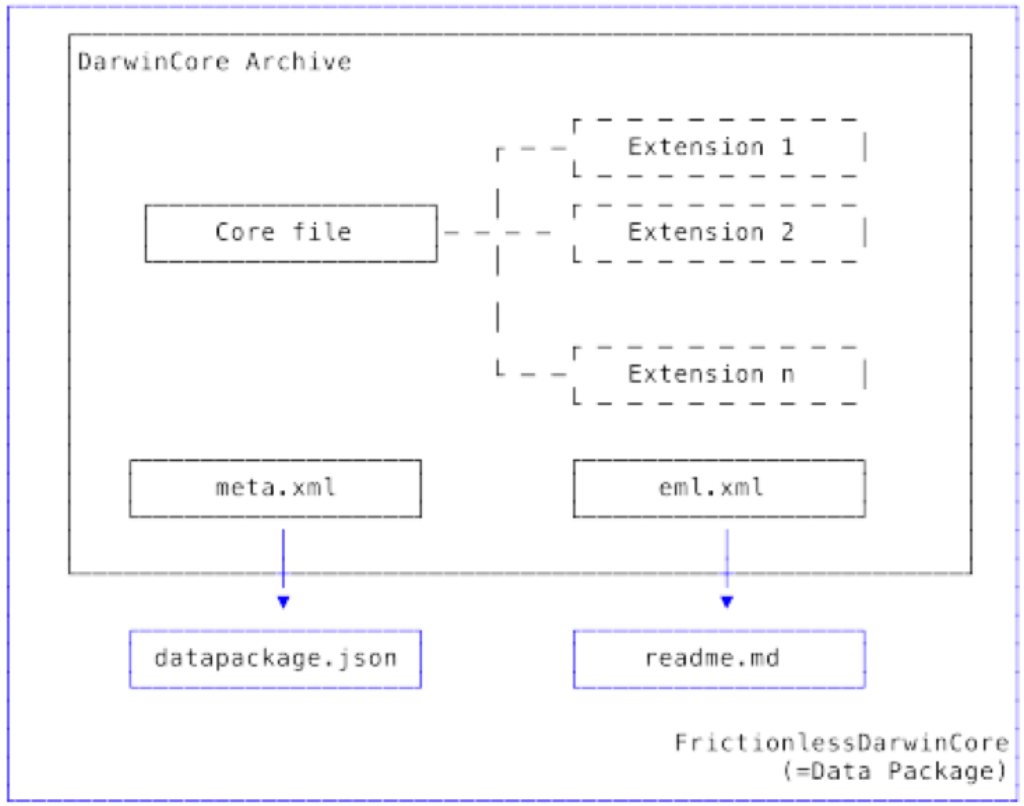


Nuevo modelo conceptual para Darwin Core

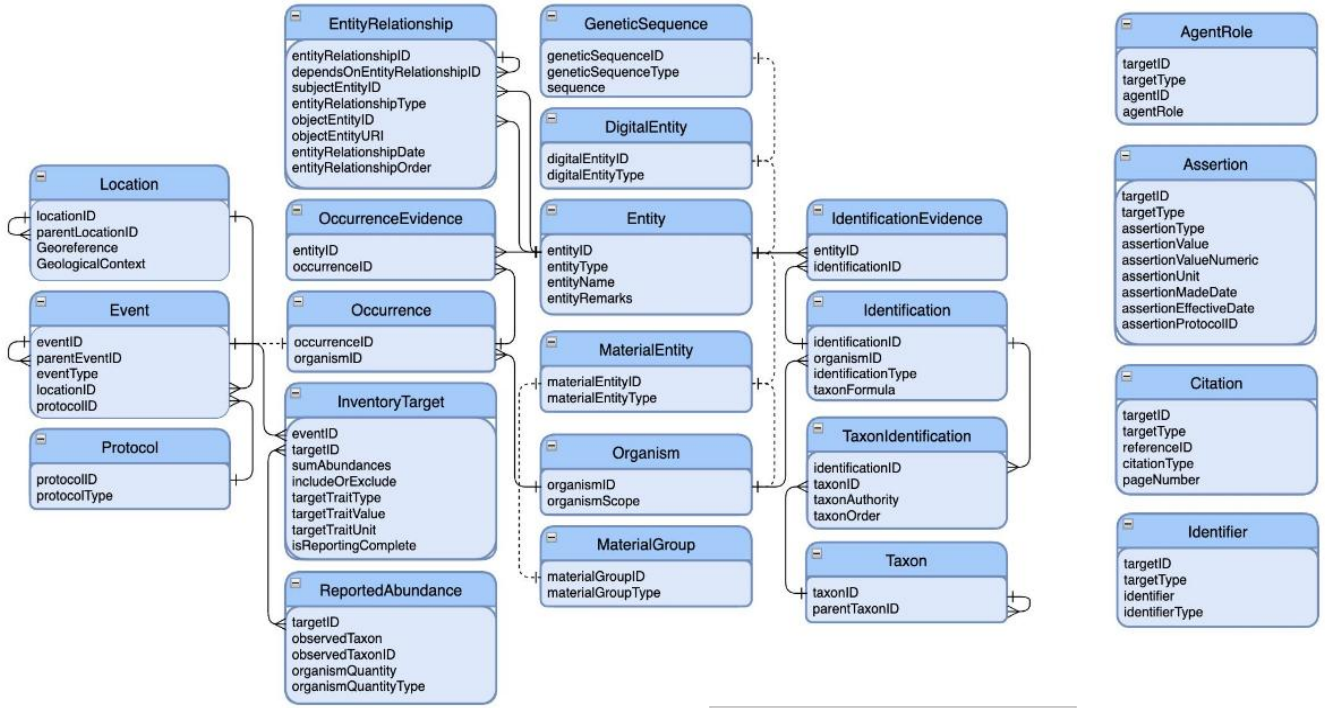


Darwin Core – DP

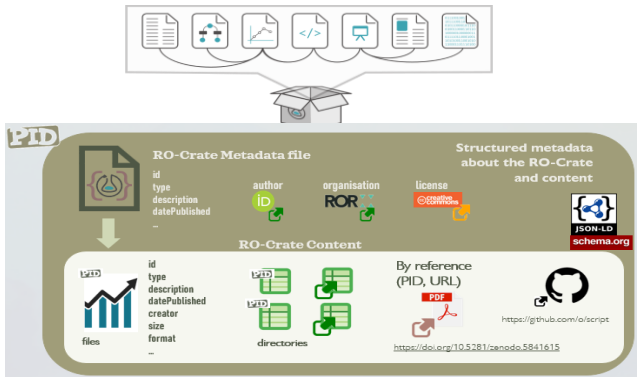




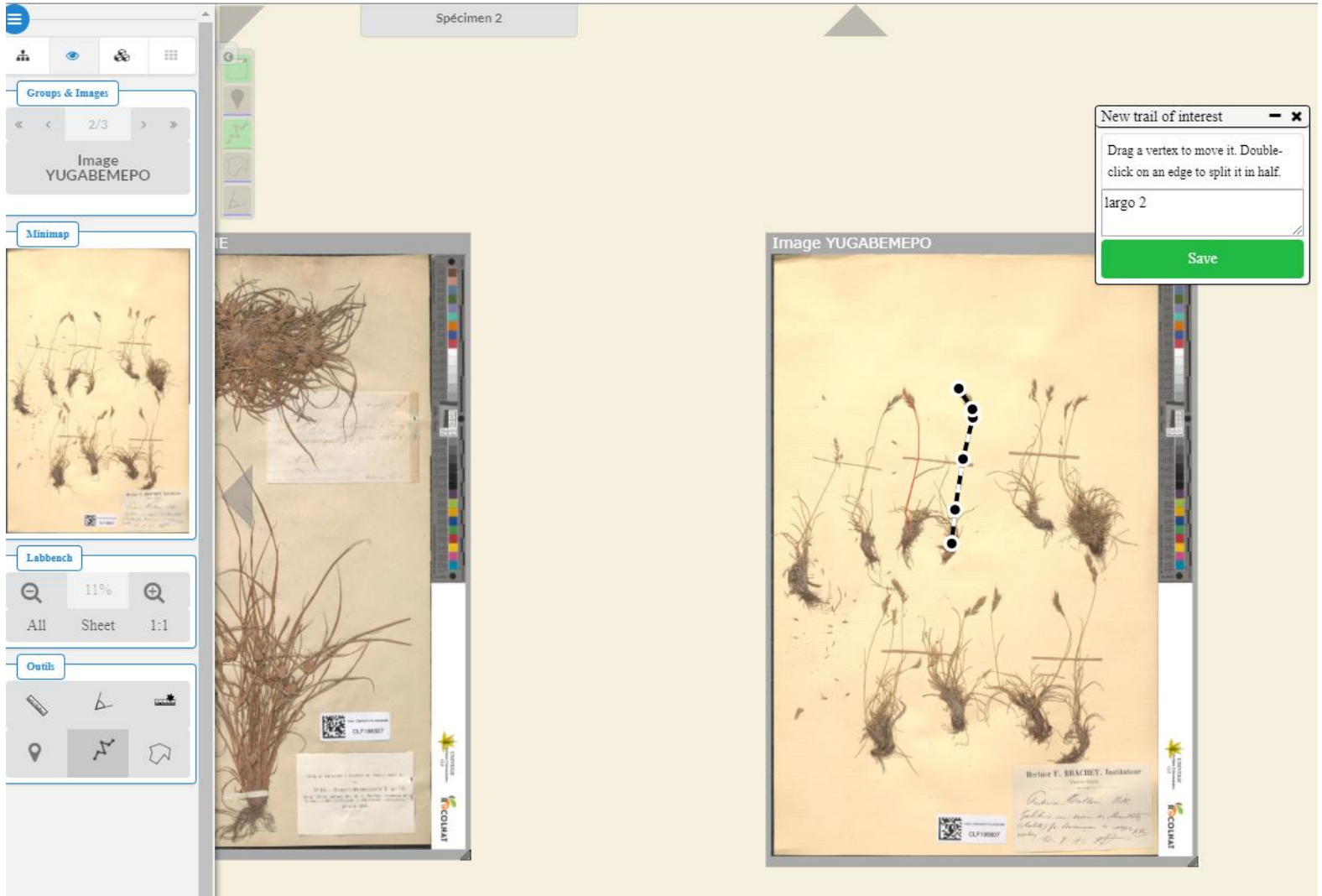
https://gbif.github.io/dwc-dp/dp/



Darwin Core – DP es una implementación de “Frictionless data Packages”



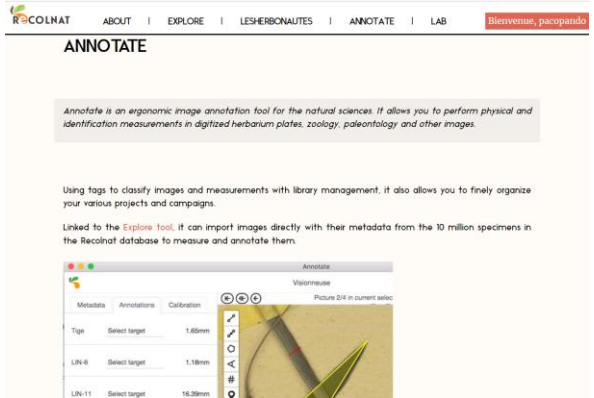
Entornos virtuales de investigación
El “Colaboratoire” de “Les Herbonautes”



La plataforma de préstamos virtuales, que permite comparar, medir y anotar ejemplares en línea y con acceso protegido

https://www.recolnat.org/en/annotate

(actualmente fuera de línea)



Herbarios virtuales

https://herbarivirtual.uib.es/

AVH
The Australasian Virtual Herbarium

Australian and New Zealand herbaria house over eight million plant, algae and fungi specimens. Herbarium specimens are an important resource for research on the Australasian flora and provide a permanent record of the occurrence of a species at a particular place and time. The AVH provides access to the collecting data associated with these specimens.

Search... Quick search

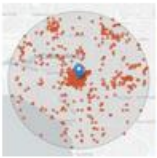
Advanced search

News

ALA Data Quality filters in AVH

In response to the community consultation that was carried out in early 2019, ALA identified data quality as one of..... Read more...

Explore your area



The 'Explore your area' functionality has been available in the ALA for years, but I for one never knew it..... Read more...

Australia's Virtual Herbarium helps unlock the history of invasive species

A team of Victorian researchers has analysed data extracted from over three million plant specimens to reconstruct the introduction of..... Read more...

https://avh.chah.org.au/

Bienvenido al Herbario Virtual de las colecciones botánicas españolas

Este portal producto de la colaboración entre la AHIM (Asociación de Herbarios Ibero-macaronésicos) y GBIF.ES (Nodo de GBIF en España), te permite explorar los ejemplares digitalizados de las colec
System of Scientific Collections) y a
y digital de colecciones europeas de

Las instituciones miembro de la AHI

Entidades

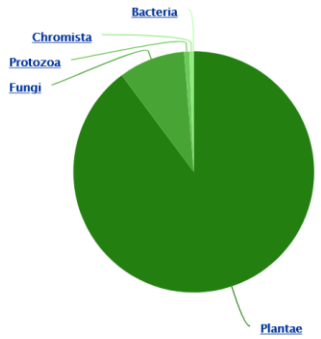
- Asociación BIGA
- Departamento Biología de Organismos y Sistemas, Universidad de Oviedo
- Departamento Biología Vegetal, Fac. Biología, Universidad de Murcia
- Departamento Ciencias Ambientales, Fac. Ciencias, Universidad de Girona
- Departamento de Biodiversidad, Ecología y Evolución, Unidad Docente de Botánica. Fac. de Ciencias Biológicas, Univ. Complutense de Madrid

Colecciones

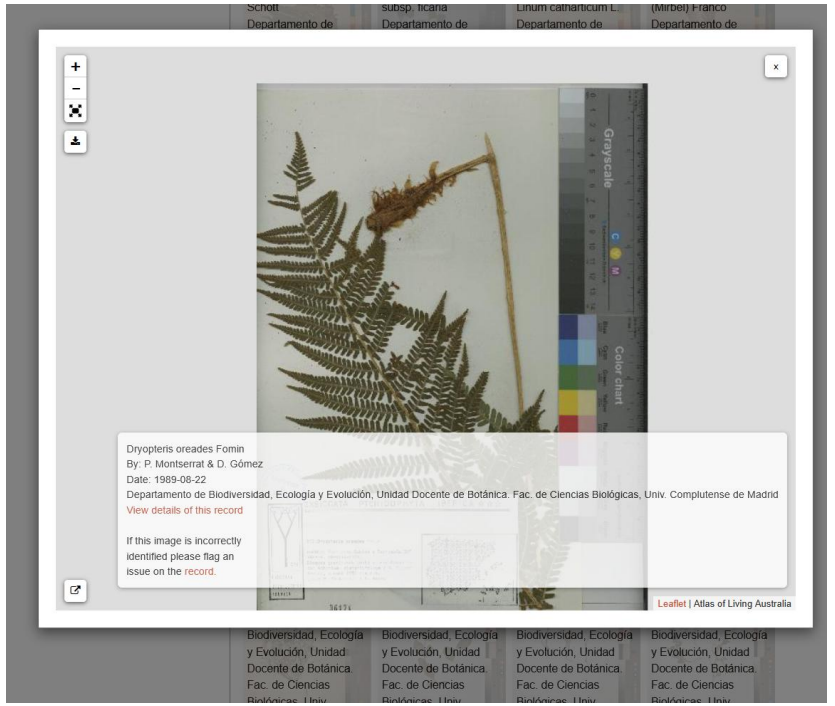
- Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB)
  - Herbario de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid (MACB)

- Departamento de Biología Vegetal y Ecología. Facultad de Ciencias Experimentales. Universidad de Almería

Taxonomías



Entidades



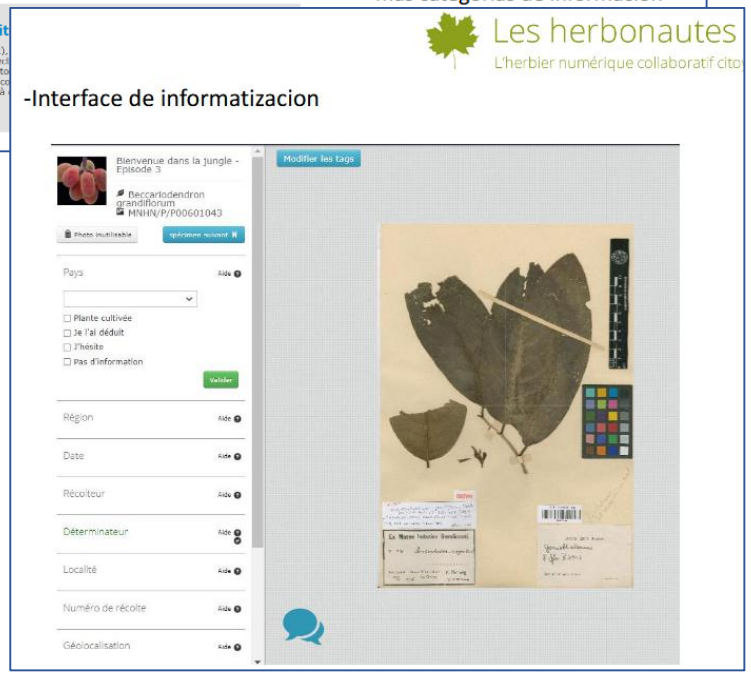
https://colecciones-botanicas.gbif.es

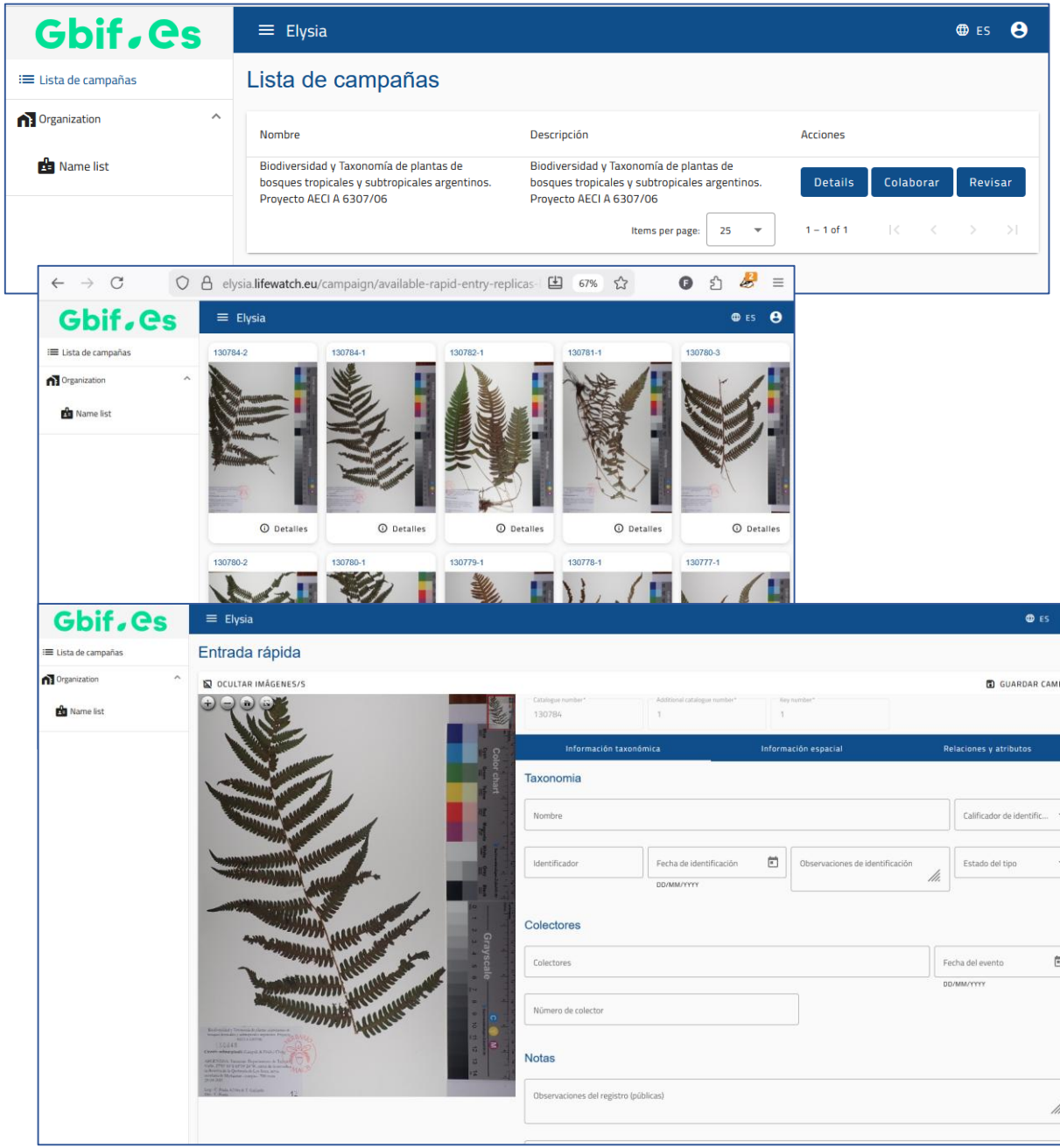
https://colecciones-botanicas.gbif.es/registros/occurrences/search?q=data_resource_uid%3Adr779#tab_mapView

Mision : lote entre 3 000/ 6 000 imagenes. Cada mision es seguida por un jefe de mision y el animador H

Cada imagen es propuesta a 2 o 3 participantes («doble ciego»)

Se proponen cuestionarios para pasar de nivel y aceder a mas categorias de informacion





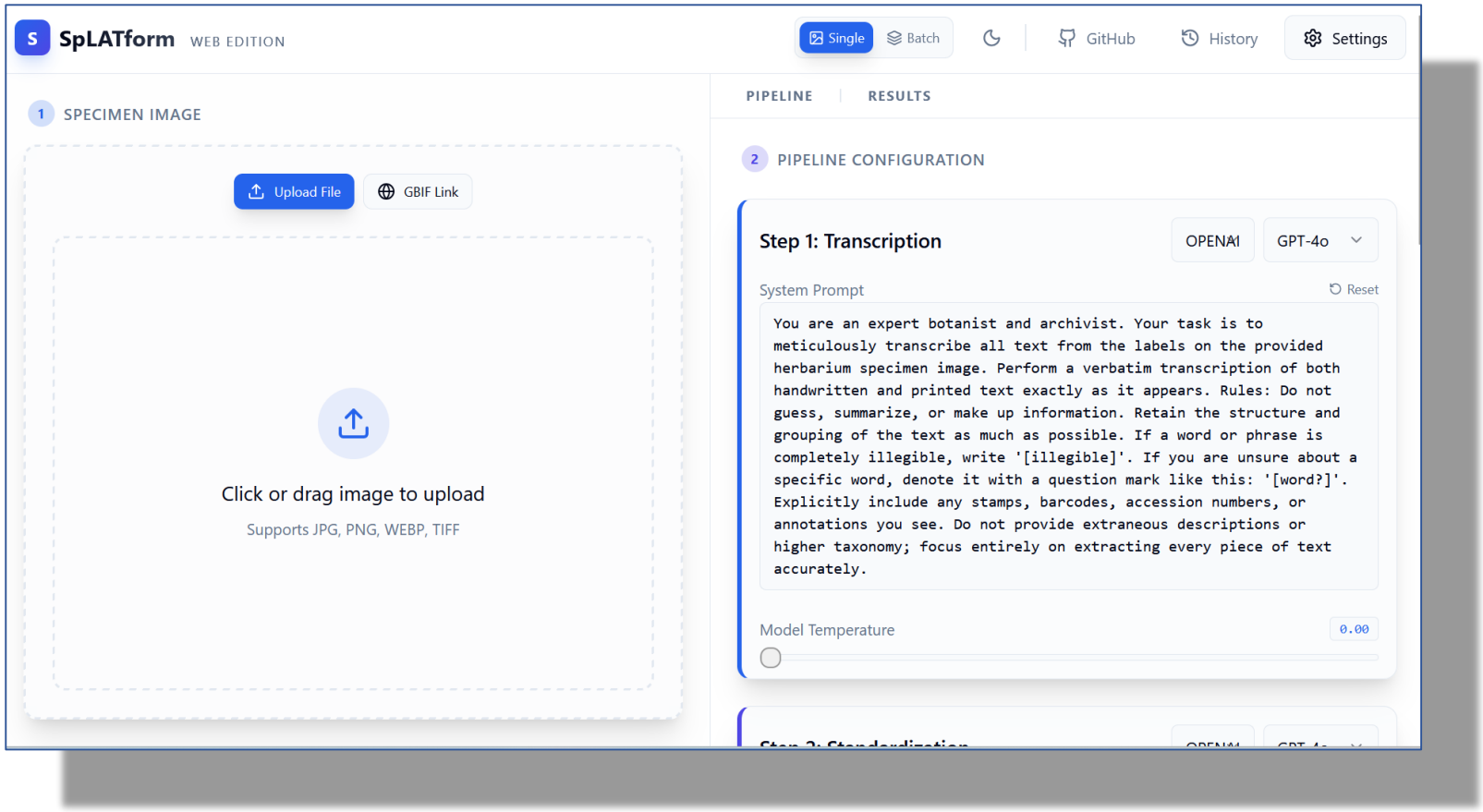
CIENCIA CIUDADANA

[Les Herbonautes, articulación entre las colecciones virtuales de herbarios y la sociedad](#) - Eva Perez Pimparé (RECOLNAT - Muséum National d'Histoire Naturelle)

https://elysia.lifewatch.eu/

IA para transcribir etiquetas

- https://gbif-norway.github.io/splatform/



Step 1: Transcription

You are an expert botanist and archivist. Your task is to meticulously transcribe all text from the labels on the provided herbarium specimen image. Perform a verbatim transcription of both handwritten and printed text exactly as it appears. Rules: Do not guess, summarize, or make up information. Retain the structure and grouping of the text as much as possible. If a word or phrase is completely illegible, write '[illegible]'. If you are unsure about a specific word, denote it with a question mark like this: '[word?]'. Explicitly include any stamps, barcodes, accession numbers, or annotations you see. Do not provide extraneous descriptions or higher taxonomy; focus entirely on extracting every piece of text accurately.

Step 2: Standardization

You are a biodiversity data steward. Your task is to structure the provided verbatim transcription of a herbarium specimen label into a valid JSON object using exclusively valid Darwin Core (DwC) terms.
Use the following Darwin Core definitions to accurately map the transcribed text:
- `dwc:scientificName`: The full scientific name on the label, including authorship if present.
- `dwc:recordedBy`: A list of names of people who collected the specimen (the collectors).
- `dwc:recordNumber`: The identifier given to the collection event by the collector (often the collector number appearing near their name).
- `dwc:eventDate`: The date the specimen was collected (format as YYYY-MM-DD if unambiguous, otherwise preserve verbatim).
- `dwc:identifiedBy`: The name(s) of the person(s) who identified the species (often marked with 'Det.').
- `dwc:country`: The country where the specimen was collected.
- `dwc:stateProvince`: The state, province, or major administrative region.
- `dwc:county`: The county or secondary administrative region.
- `dwc:locality`: The specific text describing the exact place of collection.
- `dwc:habitat`: The environment or habitat description (e.g., 'growing in sandy soil', 'oak forest').
- `dwc:verbatimElevation`: The original description of the elevation or altitude.
- `dwc:catalogNumber`: The herbarium's barcode, accession number, or physical stamp identifier.

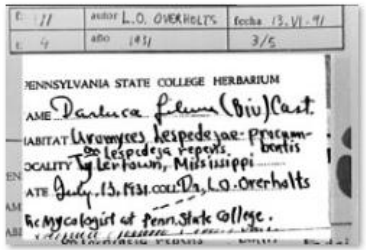
Strict constraints:
1. Output ONLY valid JSON. Do not include markdown block formatting (e.g., ```json) or any conversational text.
2. Use ONLY the 'dwc:' prefixed terms above as keys.
3. If a term has multiple values (e.g., multiple collectors for `dwc:recordedBy`), separate the values with a pipe character `|` (e.g., "John Doe | Jane Smith").
4. If the transcription lacks data for a specific DwC term, do NOT include that key in the JSON. Do not use placeholders like "Unknown" or null.
5. Do not hallucinate or add outside geographic knowledge not supported by the transcription.

You are a biodiversity data steward. Your task is to structure the provided verbatim transcription of a herbarium specimen label into a CSV file using exclusively valid Darwin Core (DwC) terms.

Use the following Darwin Core definitions to accurately map the transcribed text:
- `dwc:scientificName`: The full scientific name on the label, including authorship if present.
- `dwc:recordedBy`: A list of names of people who collected the specimen (the collectors).
- `dwc:recordNumber`: The identifier given to the collection event by the collector (often the collector number appearing near their name).
- `dwc:eventDate`: The date the specimen was collected (format as YYYY-MM-DD if unambiguous, otherwise preserve verbatim).
- `dwc:identifiedBy`: The name(s) of the person(s) who identified the species (often marked with 'Det.').
- `dwc:country`: The country where the specimen was collected.
- `dwc:stateProvince`: The state, province, or major administrative region.
- `dwc:county`: The county or secondary administrative region.
- `dwc:locality`: The specific text describing the exact place of collection.
- `dwc:habitat`: The environment or habitat description (e.g., 'growing in sandy soil', 'oak forest').
- `dwc:verbatimElevation`: The original description of the elevation or altitude.
- `dwc:catalogNumber`: The herbarium's barcode, accession number, or physical stamp identifier.

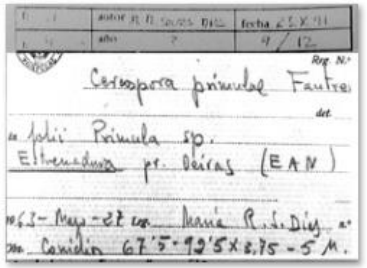
Strict constraints:
2. Use ONLY the 'dwc:' prefixed terms above as keys.
3. If a term has multiple values (e.g., multiple collectors for `dwc:recordedBy`), separate the values with a pipe character `|` (e.g., "John Doe | Jane Smith").
4. If the transcription lacks data for a specific DwC term, do NOT include that key in the JSON. Do not use placeholders like "Unknown" or null.
5. Do not hallucinate or add outside geographic knowledge not supported by the transcription.

Ejemplos de Etiquetas



3_5.jpg

4_9.jpg



9_12.jpg

20_8.jpg

label 1.jpg

label 2.jpg

label 3.jpg

label 4.jpg

label 6.jpg

Para explorar

- **ChatGPT** → https://chatgpt.com/
  Asistente conversacional versátil. Redacción, programación, análisis y resolución de problemas. Interacción natural para múltiples tareas.
- **Claude** → https://chaton.ai/claude/
  IA enfocada en razonamiento y textos largos. Buen manejo de contexto y documentos extensos. Útil para análisis y revisión detallada. Ideal para código
- **NotebookLM** → https://notebooklm.google.com/
  Herramienta centrada en documentos del usuario. Resume, organiza y responde preguntas sobre las fuentes cargadas. Ideal para estudio e investigación.

Francisco Pando
Unidad de Coordinación de GBIF.ES
CSIC
Joaquín Costa 22, 28002 Madrid

pando@gbif.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

GBIF.ES

https://mutis.rjb.csic.es