

Taller GBIF.ES: Estandarización, documentación y publicación de datos de biodiversidad en colecciones

16 – 17 de septiembre de 2026, Institut Botànic, Barcelona

Estándares para la información de la biodiversidad

Francisco Pando
GBIF.ES



**institut
botànic**

Centre mixt



Guión

- Datos y ciencia
- Ciencia abierta y estándares
- Tipos de estándares
- TDWG
- Estándares en GBIF
 - Darwin Core
 - Plinian Core
- Darwin Core Archive
 - Extensiones para colecciones
- GBIF new data model
- Darwin Core – DP



Elementos que hacen la ciencia abierta

- Metadatos
- **Estándares**
- Identificadores
- Licencias
- Citación
- El cómo se sirve el dato

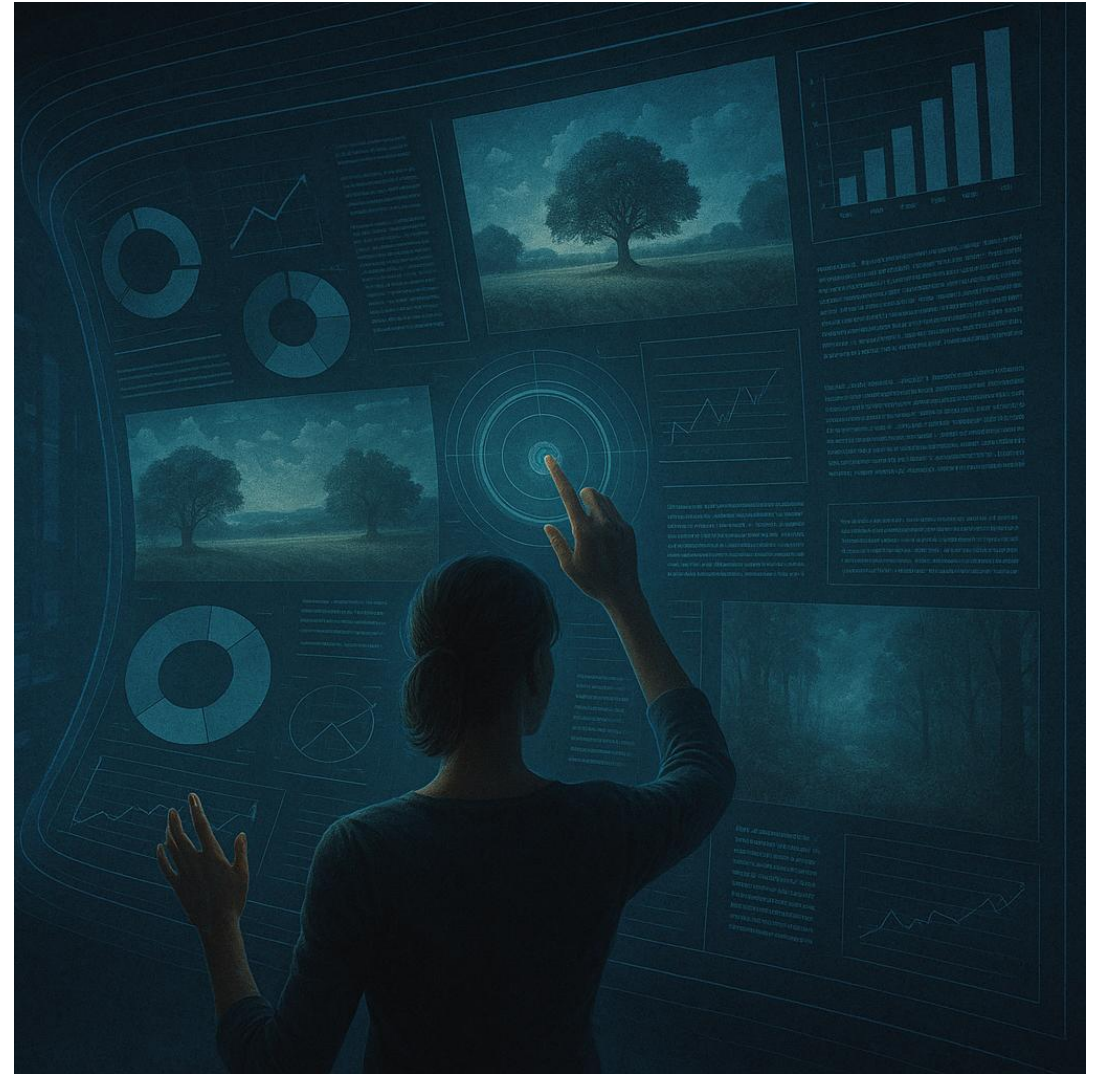
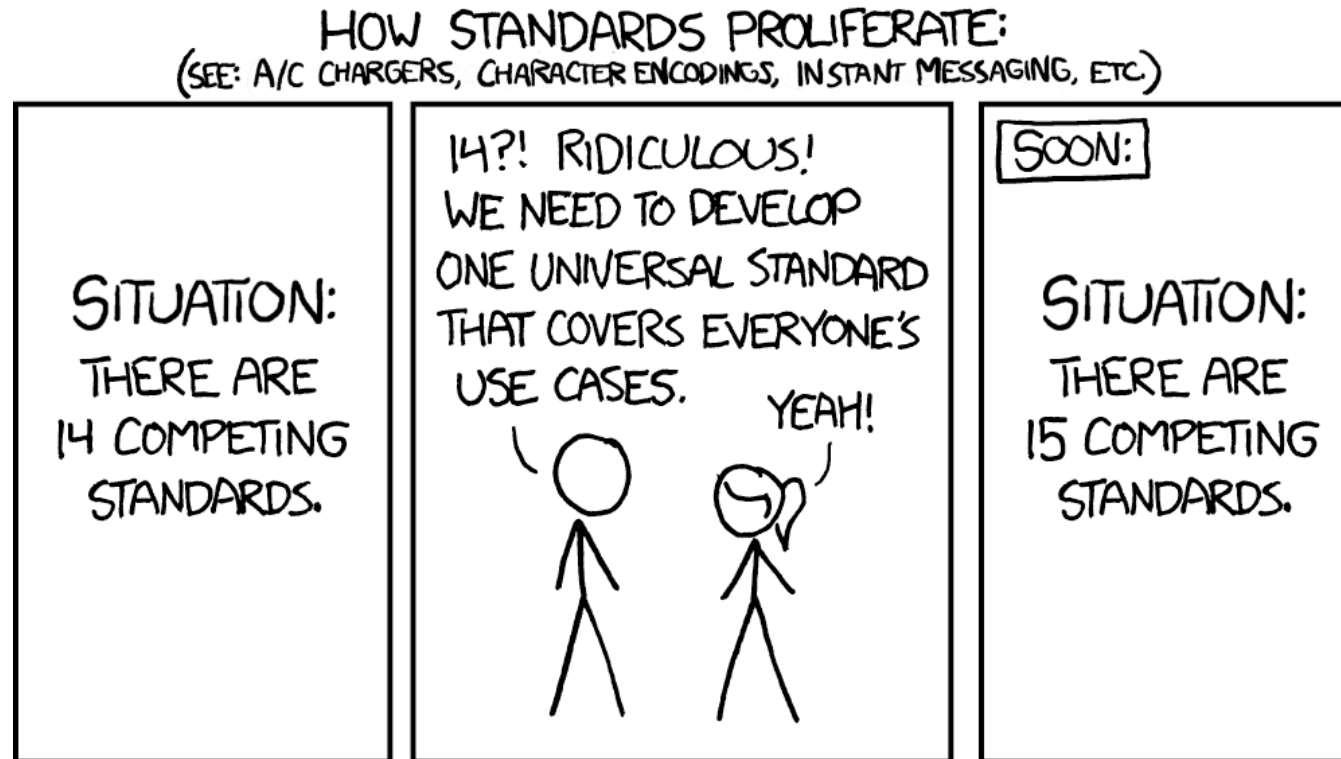


Imagen generada con
IA

Estándares



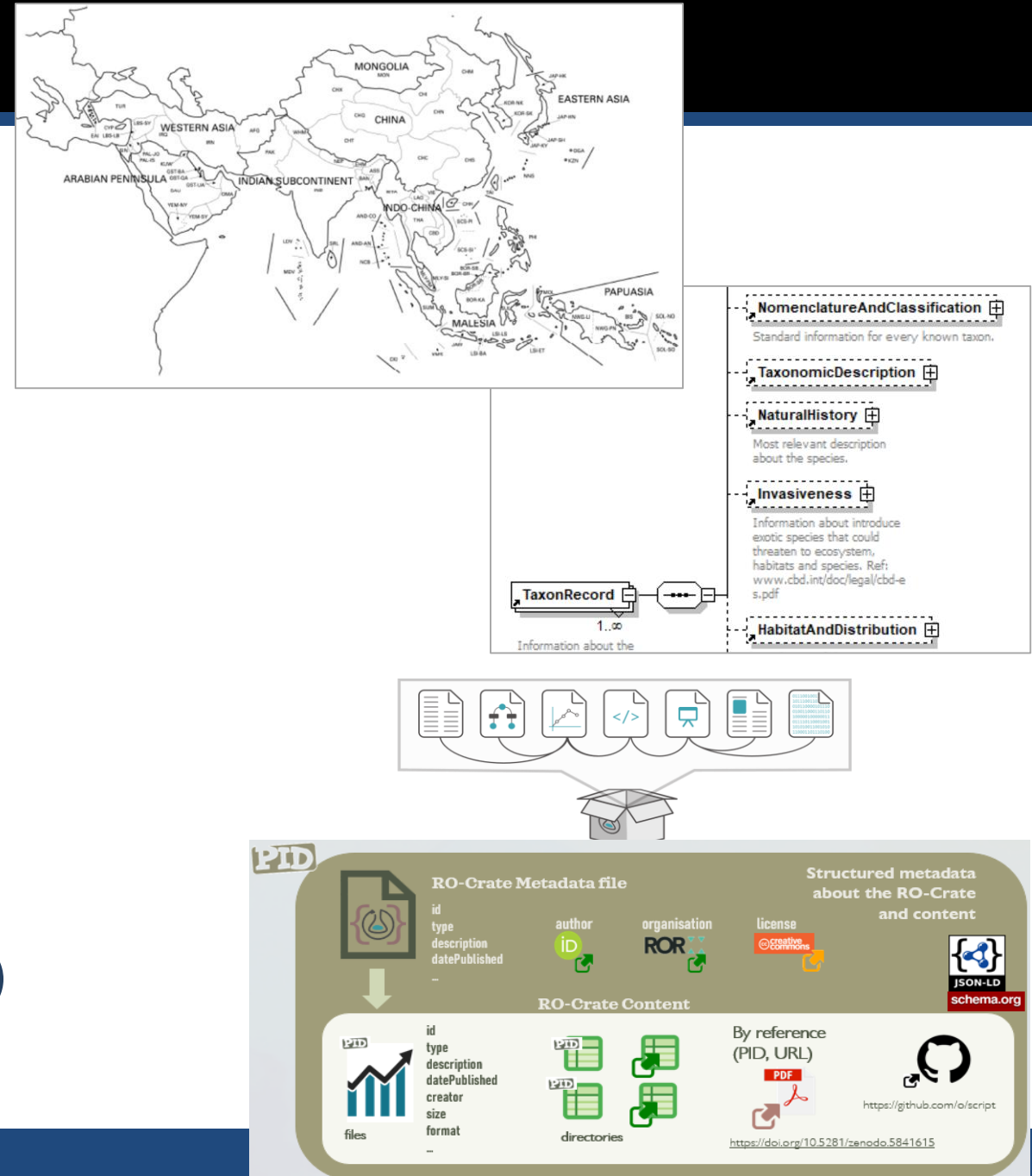
- Si lo cambias, ya no es un estándar
- Si solo lo usas tu, no es un estándar
- Los estándares son dinámicos y están, en constante evolución: revisiones y ediciones

<https://xkcd.com/927/>

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

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)



“Estándares orquestados”

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

```
Xml                                                                    Cop
<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:des
  <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
</metadata>
```

Biodiversity Information Standards (TDWG)

We are a non-profit organization and a community dedicated to
developing biodiversity information standards.

<https://www.tdwg.org>

ABCD

Access to Biological Collection Data (ABCD) Schema

Audiovisual Core

Audiovisual Core Multimedia Resources Metadata Schema

Darwin Core

Darwin Core

GUID Applicability Statements

GUID and Life Sciences Identifiers Applicability Statements

Latimer Core

Latimer Core, a standard for describing collections

SDS

TDWG Standards Documentation Standard (SDS)

TAPIR

TDWG Access Protocol for Information Retrieval (TAPIR)

TCS

Taxon Concept Schema (TCS)

VMS

Vocabulary Maintenance Standard (VMS)

WGSRPD

World Geographical Scheme for Recording Plant Distributions (WGSRPD)

Interest and Task Groups

- MG** Access to Biological Collections Data
- IG** Annotations
- IG** Attribution
 - TG** People in Biodiversity Data
- MG** Audiovisual Core
 - TG** 3D Imagery and Data
 - TG** Views Controlled Vocabularies
- IG** Biodiversity Data Quality
 - TG** Framework on data quality
 - TG** Data quality tests and assertions
 - TG** Data quality use cases
 - TG** Best practices for development of vocabularies of values ("Vocabularies")
- IG** Biodiversity Informatics Curriculum
- IG** Biodiversity Services and Clients
- IG** Biological Interactions Data
- IG** Citizen Science
- IG** Collection Descriptions
 - MG** Latimer Core
 - TG** Minimum Information about a Digital Specimen (MIDS)
 - TG** Modelling Research Expeditions
- MG** Darwin Core
 - TG** Invasive Species
 - TG** Plant Phenology
 - TG** Realm and Biome
 - TG** Sensitive Species Extension
- IG** Earth Sciences and Paleobiology
 - TG** Darwin Core Chronometric Age Extension
 - TG** Extension for Geosciences (EFG)
 - TG** Mineralogy Extension
- IG** Genomic Biodiversity
 - TG** Environmental Samples and eDNA
 - TG** Sustainable DarwinCore MixS Interoperability
- IG** Geographical Schemes
 - TG** Marine Geo Schemes Task Group
 - TG** Terrestrial GeoSchemes TG
- IG** Machine Observations
- IG** Observations & Specimens
 - TG** How Did It Die?
 - TG** Humboldt Extension
 - TG** Material Sample
- IG** Process
- IG** Species Information
 - TG** Invasive Organism Information
 - TG** Plinian Core
- MG** Taxon Concept Schema (TCS) Maintenance Group
 - TG** Taxonomic Concept Schema (TCS2) Task Group
- IG** Technical Architecture Group
 - TG** Standards Mapping Task Group

Biodiversity Information Standards (TDWG)

We are a non-profit organization and a community dedicated to
developing biodiversity information standards.

<https://www.tdwg.org>

ABCD

Access to Biological Collection Data (ABCD) Schema

Audiovisual Core

Audiovisual Core Multimedia Resources Metadata Schema

Darwin Core

Darwin Core

GUID Applicability Statements

GUID and Life Sciences Identifier Applicability Statements

Latimer Core

Latimer Core, a standard for describing collections

SDS

TDWG Standards Documentation Standard (SDS)

TAPIR

TDWG Access Protocol for Information Retrieval (TAPIR)

TCS

Taxon Concept Schema (TCS)

VMS

Vocabulary Maintenance Standard (VMS)

WGSRPD

World Geographical Scheme for Recording Plant Distributions (WGSRPD)

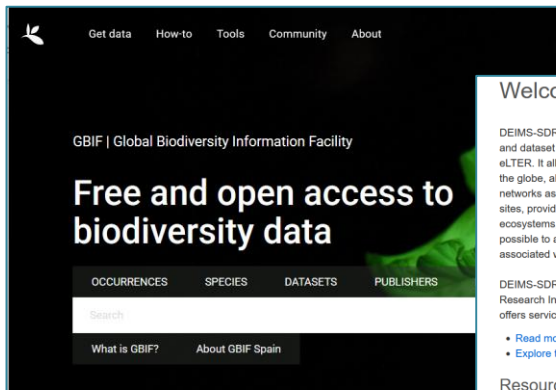
Interest and Task Groups

- MG** Access to Biological Collections Data
- IG** Annotations
- IG** Attribution
 - TG** People in Biodiversity Data
- MG** Audiovisual Core
 - TG** 3D Imagery and Data
 - TG** Views Controlled Vocabularies
- IG** Biodiversity Data Quality
 - TG** Framework on data quality
 - TG** Data quality tests and assertions
 - TG** Data quality use cases
 - TG** Best practices for development of vocabularies of values ("Vocabularies")
- IG** Biodiversity Informatics Curriculum
- IG** Biodiversity Services and Clients
- IG** Biological Interactions Data
- IG** Citizen Science
- IG** Collection Descriptions
 - MG** Latimer Core
 - TG** Minimum Information about a Digital Specimen (MIDS)
 - TG** Modelling Research Expeditions
- MG** Darwin Core
 - TG** Invasive Species
 - TG** Plant Phenology
 - TG** Realm and Biome
 - TG** Sensitive Species Extension
- IG** Earth Sciences and Paleobiology
 - TG** Darwin Core Chronometric Age Extension
 - TG** Extension for Geosciences (EFG)
 - TG** Mineralogy Extension
- IG** Genomic Biodiversity
 - TG** Environmental Samples and eDNA
 - TG** Sustainable DarwinCore MixS Interoperability
- IG** Geographical Schemes
 - TG** Marine Geo Schemes Task Group
 - TG** Terrestrial GeoSchemes TG
- IG** Machine Observations
- IG** Observations & Specimens
 - TG** How Did It Die?
 - TG** Humboldt Extension
 - TG** Material Sample
- IG** Process
- IG** Species Information
 - TG** Invasive Organism Information
 - TG** Plinian Core
- MG** Taxon Concept Schema (TCS) Maintenance Group
 - TG** Taxonomic Concept Schema (TCS2) Task Group
- IG** Technical Architecture Group
 - TG** Standards Mapping Task Group

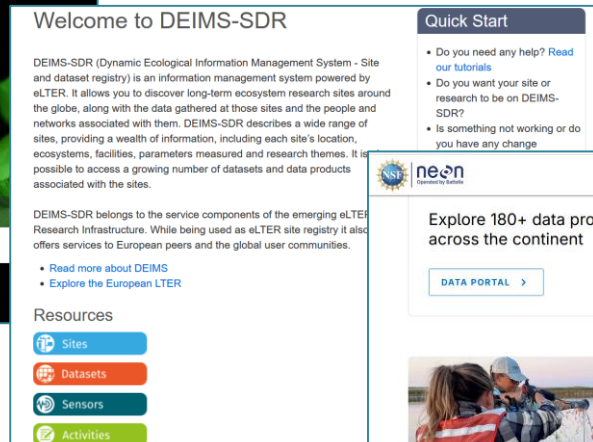
EML (Ecological Metadata Language)

>> estándar de código abierto y basado en [XML](#), diseñado para documentar datos de investigación en ciencias ambientales y ecológicas. Permite describir el contenido, alcance espacial y temporal, métodos y protocolos de los datos

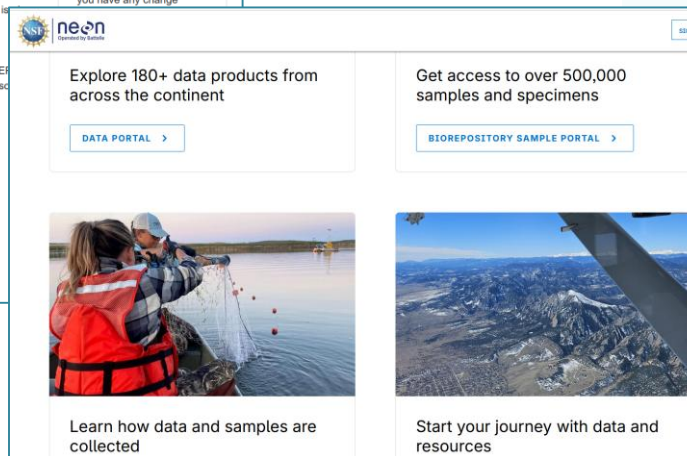
<https://www.gbif.org>



<https://deims.org>

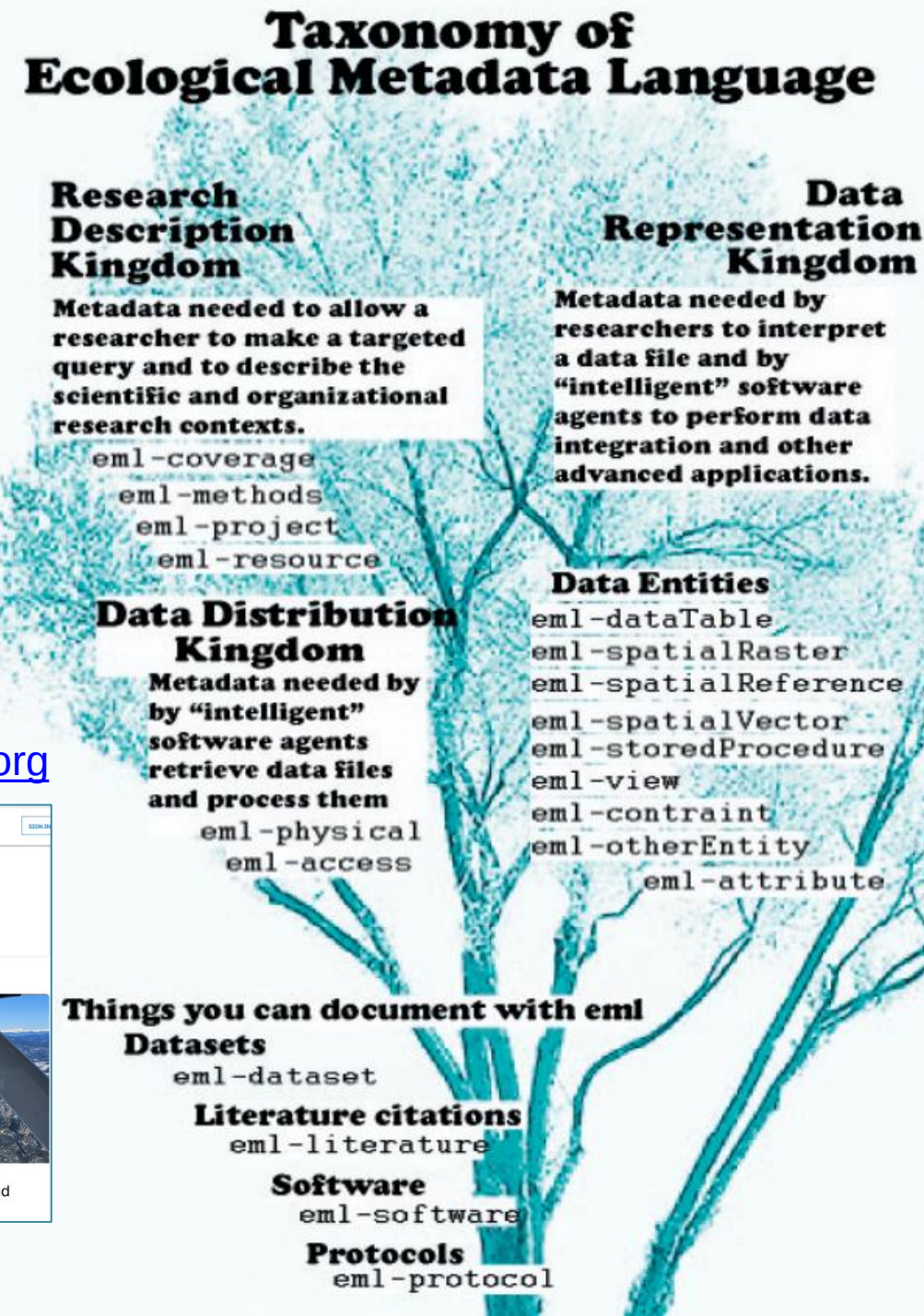


www.neonscience.org



<https://eml.ecoinformatics.org/>

<https://lternet.edu/wp-content/uploads/2010/12/EMLHandbook-2.pdf>



Darwin Core



TDWG

Darwin Core

Darwin Core Quick Reference Guide

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

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

Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	Generic Darwin Core (relational)
MeasurementOrFact	measurements, facts, characteristics, assertions, references	

https://www.researchgate.net/figure/Darwin-Core-Categories-Simple-Darwin-Core-is-comprised-of-seven-categories-of-terms_fig3_221741592

Plinian Core

- Compartir información a nivel de especies (un estándar de intercambio).
- Interoperabilidad: expresado en xml, se apoya en otros estándares
- Información = propiedades y traits relativos al taxón: descripciones, nomenclatura, estado de conservación, manejo, historia natural..
- Este proyecto ha sido liderado por:
 - El Instituto Nacional de Biodiversidad de Costa Rica (INBio, Costa Rica)
 - Nodo Español de GBIF (GBIF, España)
 - La Universidad de Granada (UG, España)
 - El Instituto Alexander von Humboldt (IAvH, Colombia)
 - La Comisión Nacional del Conocimiento y el Uso de Biodiversidad de México (CONABIO)
 - La Universidad de Sao Paulo y SiBBR Brazil)
- Auto-contenido y con capacidad de dar soporte a la integración de datos en múltiples bases de datos, con diferentes niveles de granularidad.
- Clases: 13 Genéricas, 27 Complejas y 22 Simples.
- Adoptado por el Ministerio de Agricultura..... Y Medio Ambiente de España como estándar
- Parte de los procesos de la organización “Biodiversity Information Standards” TDWG:
<http://www.tdwg.org/>



tdwg / PlinianCore



Code



Issues

5



Pull requests



Discussions



Actions



Projects



Wiki



PlinianCore

Public



Edit Pins



Unwatch

32



Fork

5



Plinian Core

Pages 236

Plinian Core is a standard oriented to share species level information. Its hierarchical schema allows --for instance-- to develop species pages or data sheets in websites.

[About Plinian Core](#)

Abstract Model [Current v3.2.2.7]

Available at: [xsd abstract-model stable-version](#)

In order to validate the abstract model schema, some additional schemas are needed. These are also available at the same location.

Application profiles already defined:

It's important to know how we can deploy an Application Profile. If we want to check an AP, we have to download, in the same folder where we have our AP, all the files that are located in this link: [xsd abstract-model stable-version](#). Then, when we open the AP in the editor, e.g: Altova, this one will validate the AP

Plinian Core terms documentation



- [Home](#)
- [Plinian Core Terms Index](#)
- [Plinian Core Terms arranged hierarchically](#)
 - [Metadata](#)
 - [TaxonRecord](#)
 - [BaseElements](#)
 - [RecordMetadata](#)
 - [NomenclatureAndClassification](#)
 - [TaxonomicDescription](#)
 - [NaturalHistory](#)
 - [Invasiveness](#)
 - [HabitatAndDistribution](#)
 - [DemographyAndThreat](#)

<https://github.com/tdwg/PlinianCore>

Términos

Term index

This quick guide provides a list of all current elements of the Plinian core. elements are organized by categories (in bold) in the index. The categories correspond to Plinian Core elements that are classes (elements that have elements to describe them). The elements that describe a given class (the properties) appear in the list immediately below the name of the category. The index provides links to the element descriptions in the table below the

Metadata

[Dataset_ID](#) | [dateStamp](#) | [citation](#) | [eml-dataset](#) | [resourceLogoUrl](#) | [Refer](#)

Base Elements

[TaxonRecordID](#) | [taxonConceptID](#) | [GlobalUniqueIdentifier](#) | [Abstract](#)

Record Metadata

[Language](#)
([TargetAudiences](#)) > [Audience](#) | [AudiencesUnstructured](#)
([Version](#)) > [Major](#) | [Minor](#) | [Modifier](#) | [DateIssued](#) | [PreferredFlag](#)
([Revision](#)) > [associatedParty](#) | [pubDate](#) | [created](#)

Nomenclature and Classification

([TaxonRecordName](#)) > [ScientificName](#)
([Synonyms](#)) > [SynonymName](#) | [SynonymStatus](#) | [SynonymsUnstructured](#)
([CommonNames](#)) > [Name](#) | [Language](#) | [UsedIn](#) | [UsedBy](#) | [CommonNamesUnstructured](#)
([Hierarchy](#)) > [kingdom](#) | [phylum](#) | [class](#) | [order](#) | [family](#) | [genus](#) | [subgenus](#) | [specificEpithet](#) | [infraspecificEpithet](#) | [higherClassification](#) | [ParentTaxon](#)
([MiscDetails](#)) > [MeasurementOrFact](#) | [DetailUnstructured](#)
[NomenclatureAndClassificationUnstructured](#)

Taxonomic Description

[BriefDescription](#)
([FullDescription](#)) > [MeasurementOrFact](#) | [FullDescriptionUnstructured](#)
([IdentificationKeys](#)) > [Keys](#)
[TaxonomicDescriptionUnstructured](#)

Natural History

([LifeForm](#)) > [MeasurementOrFact](#) | [LifeFormUnstructured](#)
([LifeCycle](#)) > [MeasurementOrFact](#) | [LifeCycleUnstructured](#)
([Reproduction](#)) > [MeasurementOrFact](#) | [ReproductionUnstructured](#)
([AnnualCycles](#)) > [Event](#) | [starTimeInterval](#) | [endTimeInterval](#) | [AnnualCyclesUnstructured](#)
([Feeding](#)) > [Strategy](#) | [StrategyRemarks](#) | [FeedingUnstructured](#)
([Dispersal](#)) > [Purpose](#) | [Type](#) | [StructureDispersed](#) | [Distance](#) | [DispersalUnstructured](#)
([Behavior](#)) > [MeasurementOrFact](#) | [BehaviorUnstructured](#)
([Interactions](#)) > [InteractionSpecies](#) | [InteractionSpeciesType](#) | [InteractionUnstructured](#)
([MolecularData](#)) > [MeasurementOrFact](#) | [MolecularDataUnstructured](#)
([Migratory](#)) > [Causes](#) | [Patterns](#) | [Routes](#) | [Season](#) | [MigratoryUnstructured](#)
([EcologicalSignificance](#)) > [MeasurementOrFact](#) | [EcologicalSignificanceUnstructured](#)
([MiscDetails](#))
([EnvironmentalEnvelope](#)) > [MeasurementOrFact](#) | [EnvironmentalEnvelopeUnstructured](#)
[NaturalHistoryUnstructured](#)

Invasiveness

[origin](#) | [presence](#) | [persistence](#) | [distribution](#) | [harmful](#) | [modified](#) | [startValidDate](#) | [endValidDate](#) | [countryCode](#) | [stateProvince](#) | [county](#) | [localityName](#) | [language](#) | [citation](#) | [abundance](#) | [trend](#) | [rateOfSpread](#) | [regulatoryListing](#) | [memo](#) | [publicationDate](#) | [localityType](#) | [WhatImpact](#) | [vector](#) | [Route](#) | [impactTarget](#) | [impactMechanism](#)
[InvasivenessUnstructured](#)

Habitat and Distribution

([Habitats](#)) > [MeasurementOrFact](#) | [HabitatUnstructured](#)
([Distribution](#)) > [DistributionScope](#) | [temporalCoverage](#) | [GeographicEntity](#) | [country](#) | [stateProvince](#) | [county](#) | [municipality](#) | [locality](#) | [DistributionUnstructured](#)
([Endemic](#)) > [EndemicTo](#) | [EndemicIn](#) | [EndemicUnstructured](#)
[HabitatAndDistributionUnstructured](#)

Demography and Threat

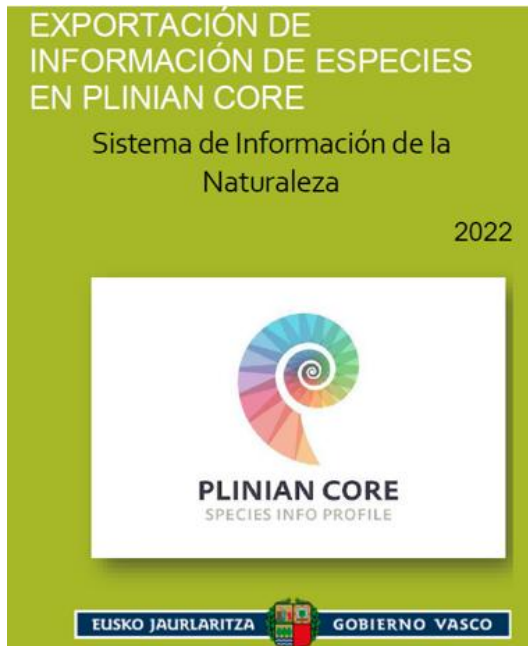
([Territory](#)) > [ExtentOfOccurrence](#) | [AreaOfOccupancy](#) | [TerritoryUnstructured](#)
([PopulationBiology](#)) > [AbundanceData](#) | [DensityData](#) | [PatternDistribution](#) | [Size](#) | [SexRatio](#) | [Fecundity](#) | [MortalityRate](#) | [BirthRate](#) | [NumberIndividualsPerObservation](#) | [AverageDensity](#) | [PopulationTrend](#) | [Recruitment](#) | [PopulationGrowthRate](#) | [Emigration](#) | [Immigration](#) | [DescriptionLifeStages](#) | [ProportionIndividualsPerStageLife](#) | [CarryingCapacity](#) | [PopulationBiologyUnstructured](#)
([ThreatStatus](#)) > [ThreatCategory](#) | [Authority](#) | [AppliesTo](#) | [ThreatStatusUnstructured](#)
([DirectThreats](#)) > [MeasurementOrFact](#) | [DirectThreatUnstructured](#)
([Legislation](#)) > [LegislationName](#) | [ProtectionLegalStatus](#) | [LegislationRead](#) | [Status](#) | [Type](#) | [Norm](#) | [AppliesTo](#) | [LegislationUnstructured](#)
[DemographyAndThreatUnstructured](#)

Uses, Management and Conservation

([Uses](#)) > [SourceOfInformation](#) | [Use-Value](#) | [PartUsed](#) | [Users](#) | [Organisms](#) | [VernacularNameUseAnnotations](#) | [ProductionDetails](#) | [MeansOfApplication-Administration](#) | [SeasonOfAvailability-Use](#) | [Conservation-ExploitationData](#) | [UseTypeAtomized](#) | [Economics](#) | [RatingPopularity](#) | [Properties](#) | [Potential](#) | [UseNotes](#) | [UsesUnstructured](#)
([ManagementAndConservation](#)) > [Objetives](#) | [ManagementPlan](#) | [Actions](#) | [HumanAndEnvironmentalrelevance](#) | [ManagementAndConservationUnstructured](#)
[UsesManagementAndConservationUnstructured](#)

[associatedParty](#), [MeasurementOrFact](#), [References](#), [AncillaryData](#)

Plinian Core is used



Bienvenidos al inventario de especies silvestres

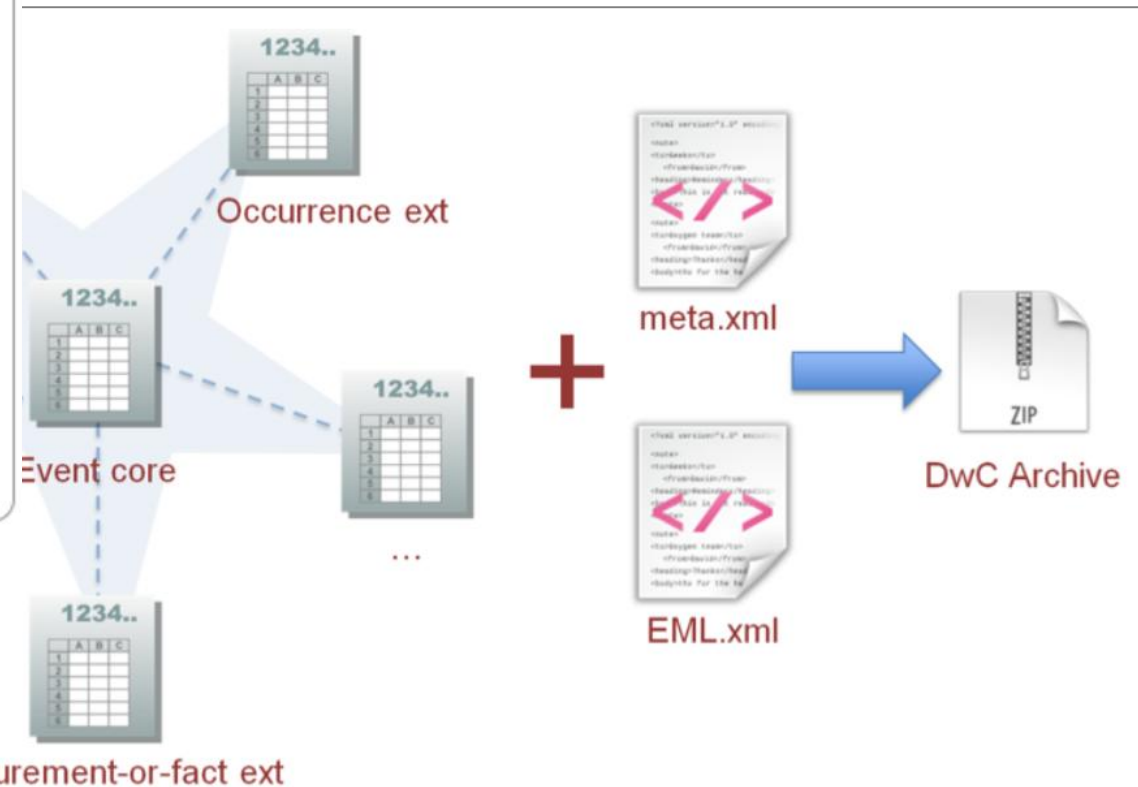
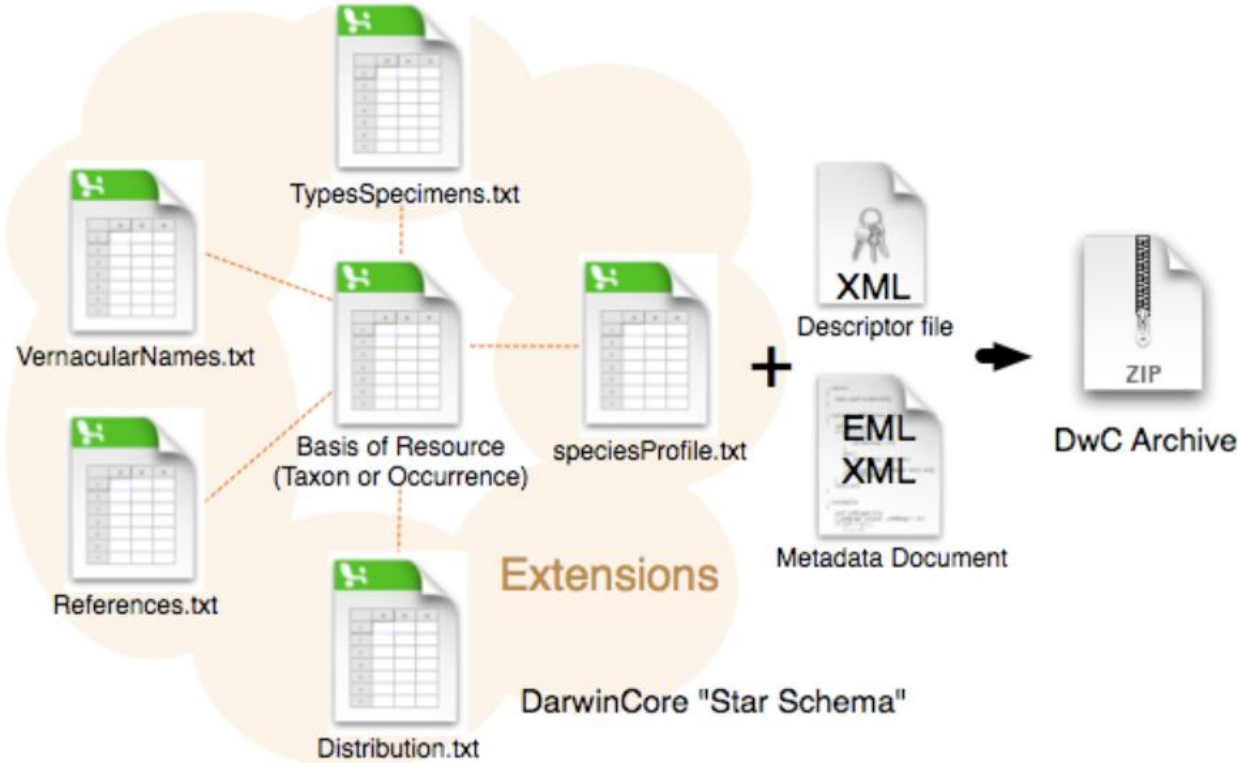
Esta página Web posee información tanto de historia natural como de gestión sobre especies silvestres nativas de Chile así como también algunas especies exóticas asilvestradas en nuestro territorio.

El Inventario Nacional de Especies cuenta solamente con un subconjunto del total de especies nativas descritas para nuestro país. Lo mismo para especies exóticas. Esperamos que al mediano plazo contemos con la totalidad de la flora y fauna nativa incorporada en este sitio.

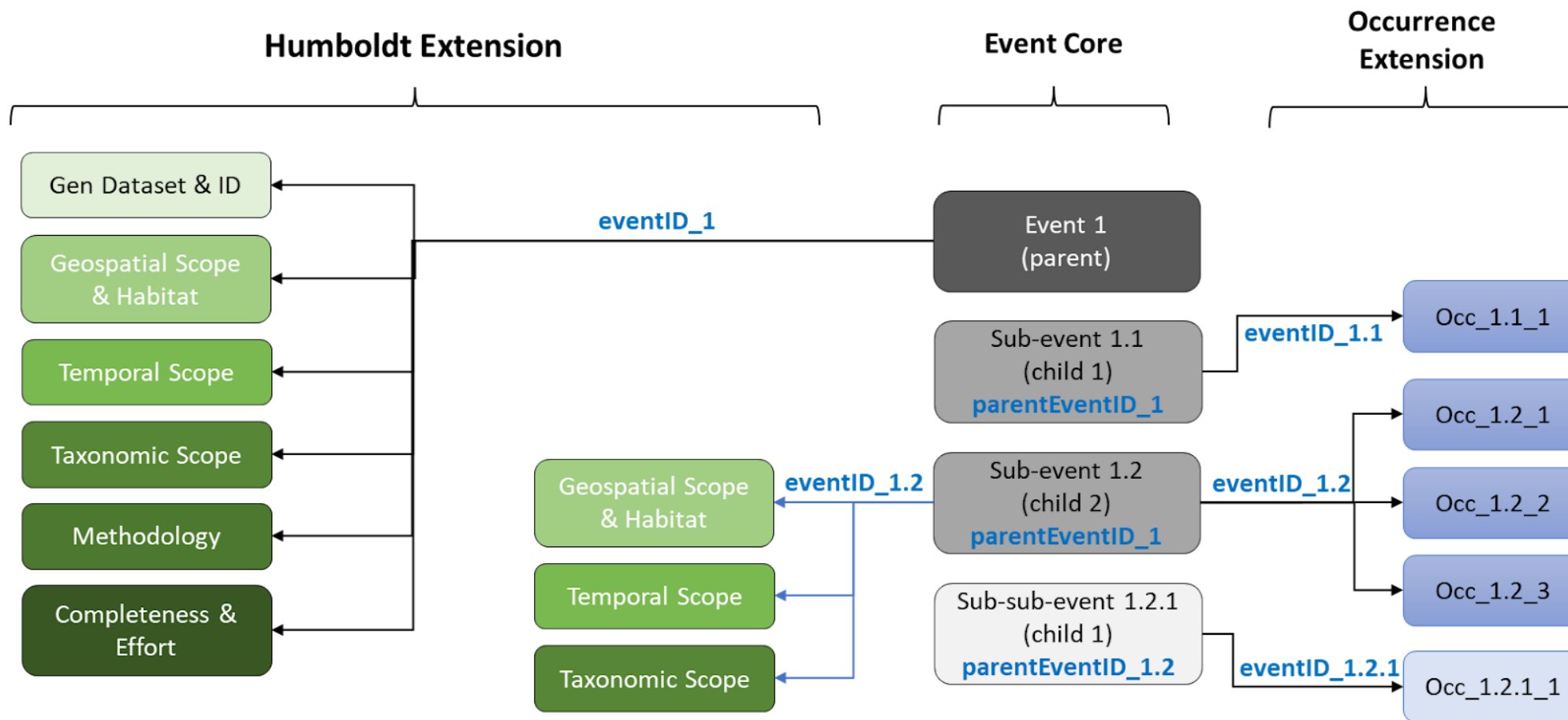
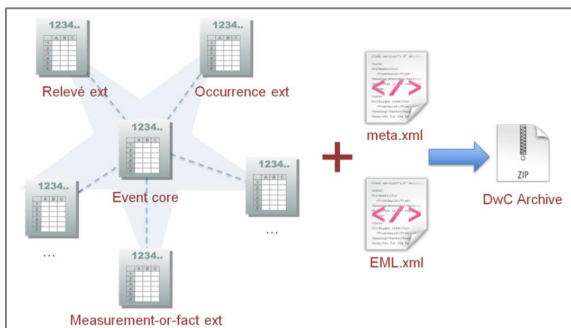


Darwin Core. Implementación: Darwin Core Archive

Darwin Core Archive Components

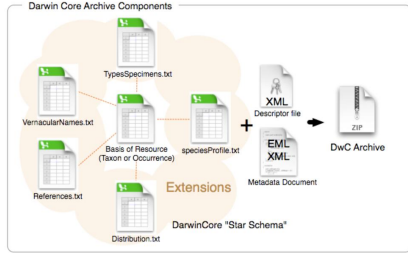


Darwin Core Archive



Visual representation of a structured biodiversity inventory using the Darwin Core Event class and the Occurrence and Humboldt Extensions. Information on different hierarchies is linked via each unique `dwc:eventID`

Extensiones de Darwin Core Archive para colecciones



Un Darwin Core Archive no se limita al núcleo de ocurrencias: puede incluir tablas de extensión repetibles, vinculadas a cada registro mediante su identificador. Así, un mismo ejemplar puede tener varias fotografías, identificaciones o medidas sin duplicar el registro principal.

Identification History

Guarda todas las identificaciones de un ejemplar a lo largo del tiempo: la vigente y todas las anteriores.

“Occurrences”

(núcleo / core)

(Especímenes, observaciones, citas)

Simple Multimedia

Enlaza cada registro con tantas imágenes, sonidos o vídeos como haga falta: pliego, etiqueta, detalle...

También a tener en cuenta en algunos casos:

Measurement or Facts

Medidas o hechos asociados al ejemplar o al evento (longitud, peso, nº de cromosomas...).

Resource Relationship

Relaciones entre registros: partes de un mismo ejemplar, duplicados, especímenes asociados...

Alternative Identifiers

Otros identificadores del mismo registro: catálogos antiguos, otros GUID, códigos de barras...

Literature References

Bibliografía asociada: la publicación en la que se citó o se basó una identificación.

Extensión: Identification History

Cada ejemplar de una colección puede recibir varias identificaciones a lo largo de su vida útil: la determinación original, revisiones posteriores, cambios de nomenclatura o verificaciones por especialistas. Esta extensión registra cada identificación como un evento independiente y fechado, sin sobrescribir ni perder las anteriores — igual que las etiquetas de redeterminación que se añaden a un pliego físico.

Campos clave

identificationID — identificador único de cada identificación.

dateIdentified — fecha en la que se realizó.

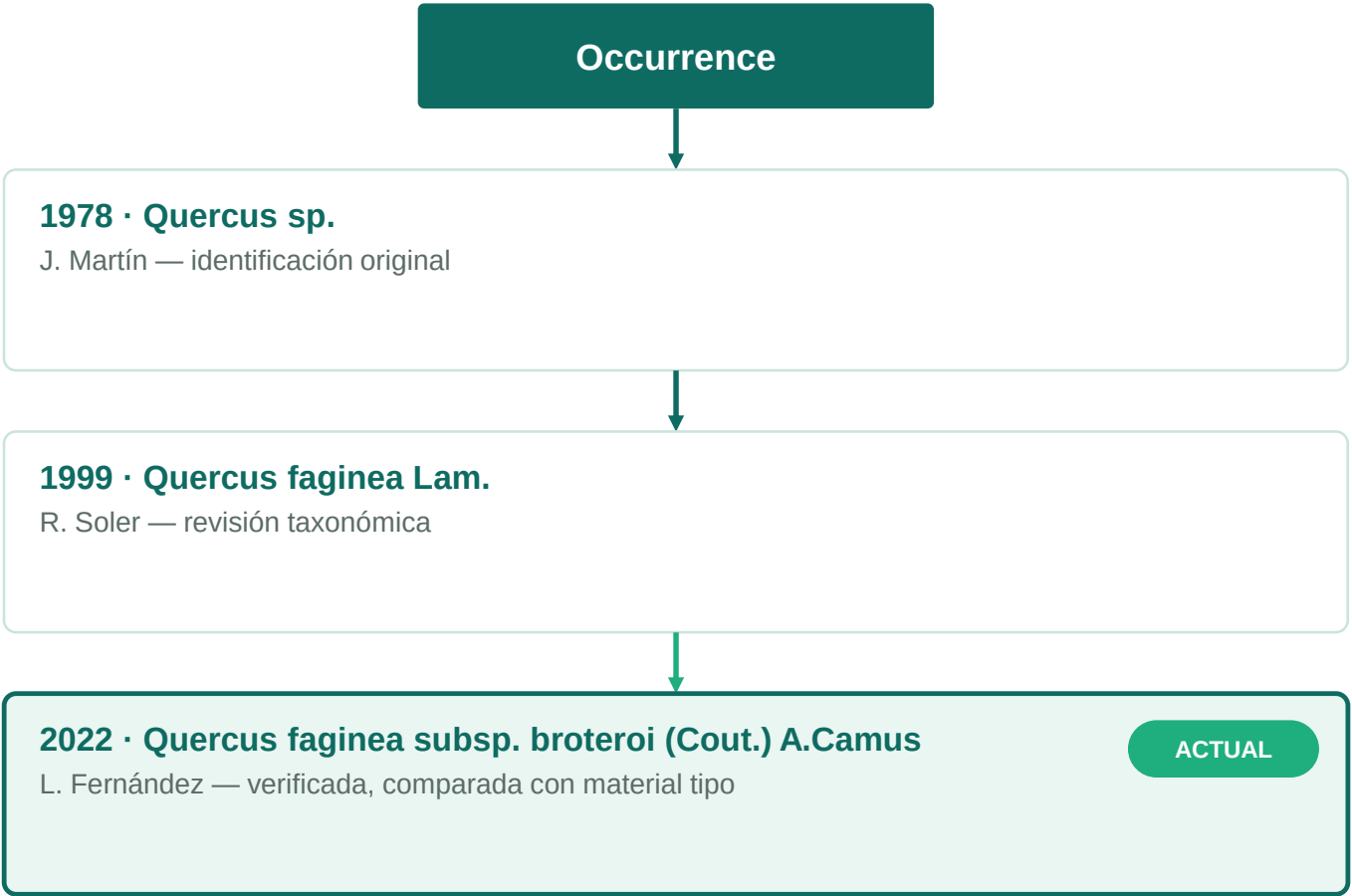
identifiedBy / identifiedByID — persona que la realizó y, si existe, su identificador (p. ej. ORCID).

typeStatus — indica si el ejemplar es un tipo nomenclatural (holotipo, isotipo...).

identificationVerificationStatus — grado de confianza o verificación de la identificación.

identificationRemarks — notas o justificación del cambio de determinación.

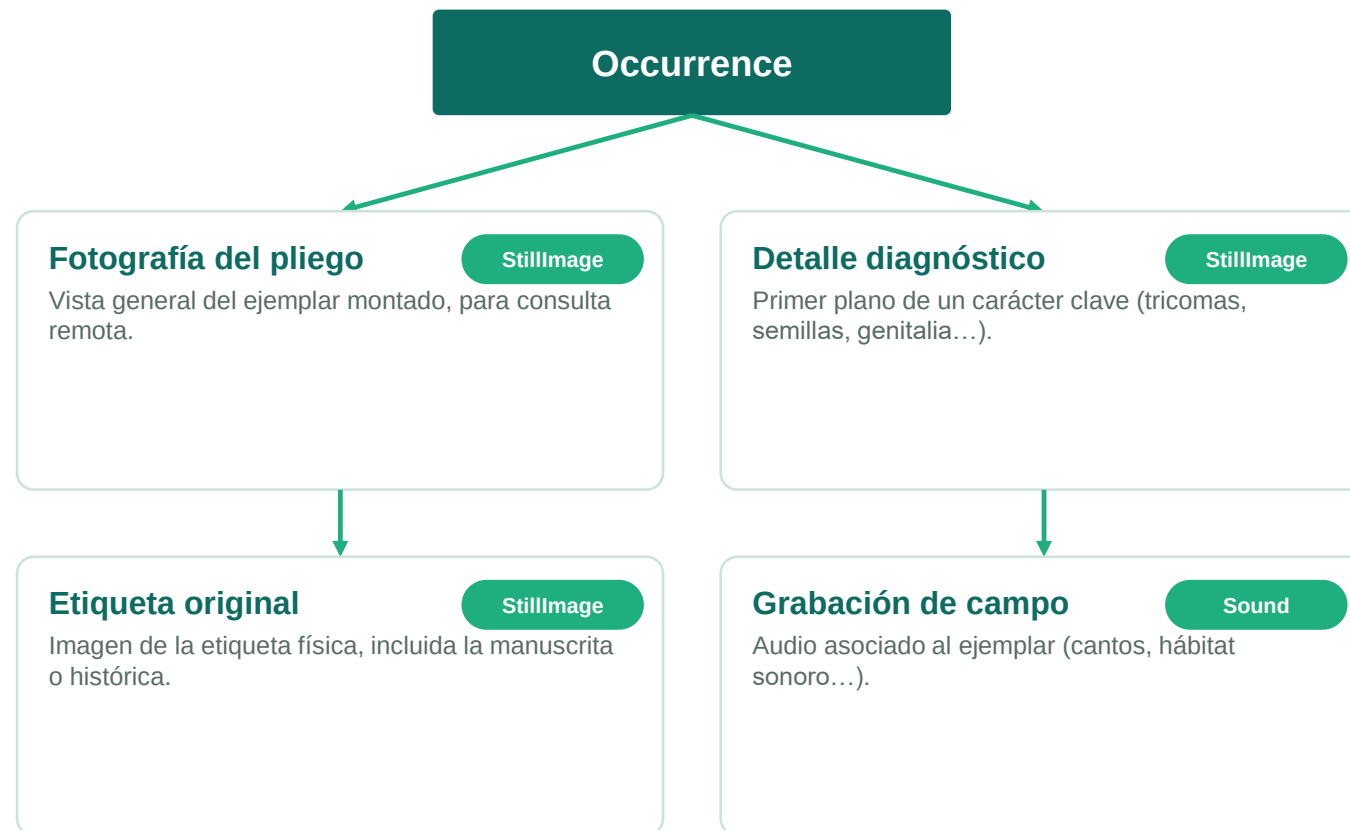
Cronología de identificaciones de un mismo ejemplar



Extensión: Simple Multimedia

Un mismo ejemplar puede tener asociados tantos archivos multimedia como sea necesario: fotografías del pliego completo, de la etiqueta, detalles diagnósticos, o incluso audio y vídeo — todo enlazado al mismo registro sin duplicarlo.

Varios recursos multimedia enlazados a un mismo registro



Campos clave

type — categoría del recurso: StillImage, Sound o MovingImage.

identifier — URL donde se aloja el archivo.

format — tipo MIME del archivo (image/jpeg, audio/mpeg...).

creator — autor o fotógrafo del recurso.

license — licencia de uso (CC BY, CC0...).

rightsHolder — titular de los derechos sobre el recurso.

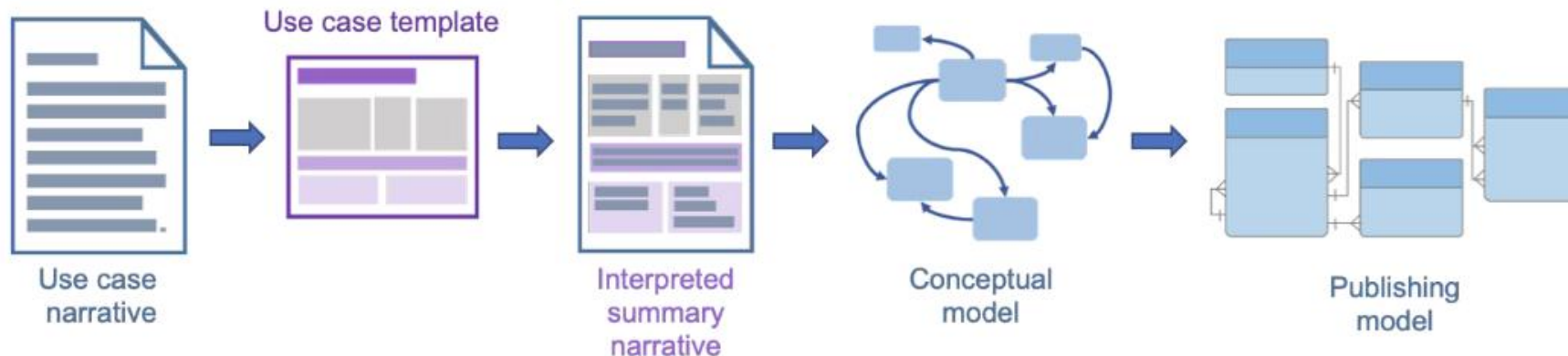
Nuevo modelo conceptual para Darwin Core

>> 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.

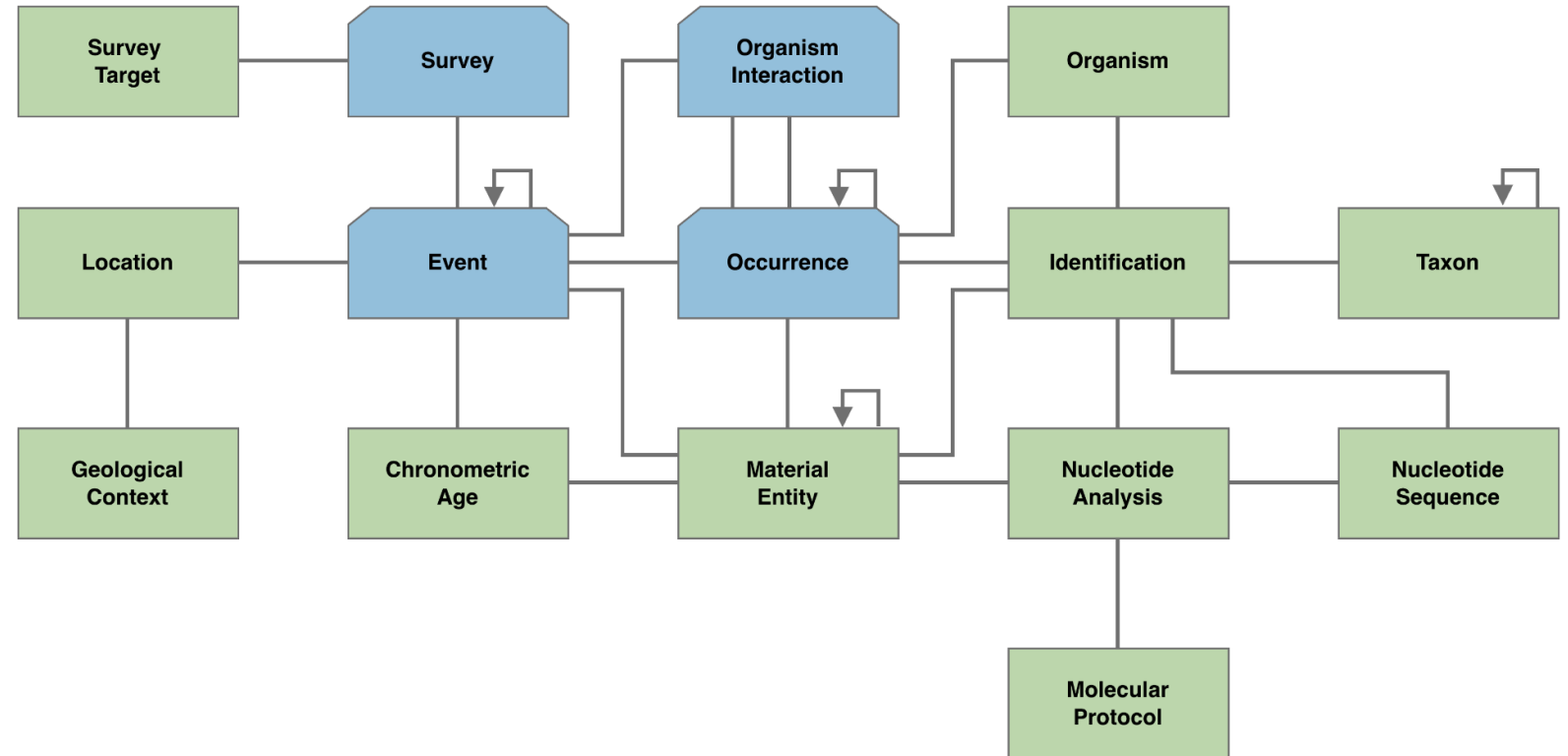
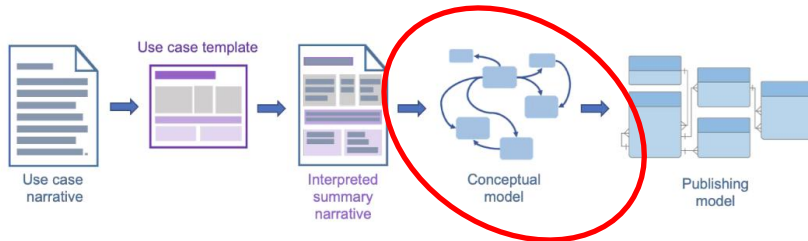
This implies a need for richer and more varied types of data than GBIF has thus far supported.”

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

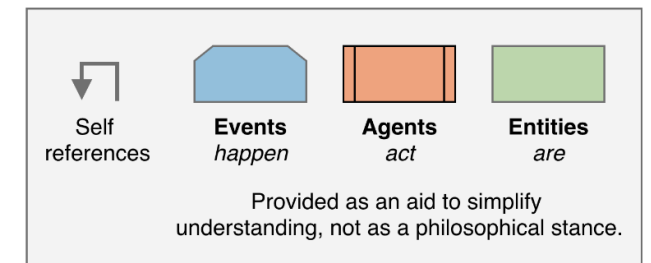
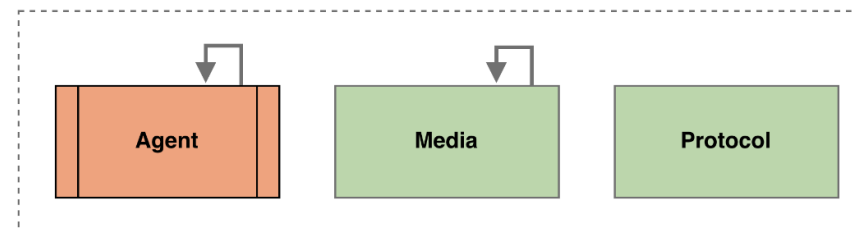


Nuevo modelo conceptual para Darwin Core

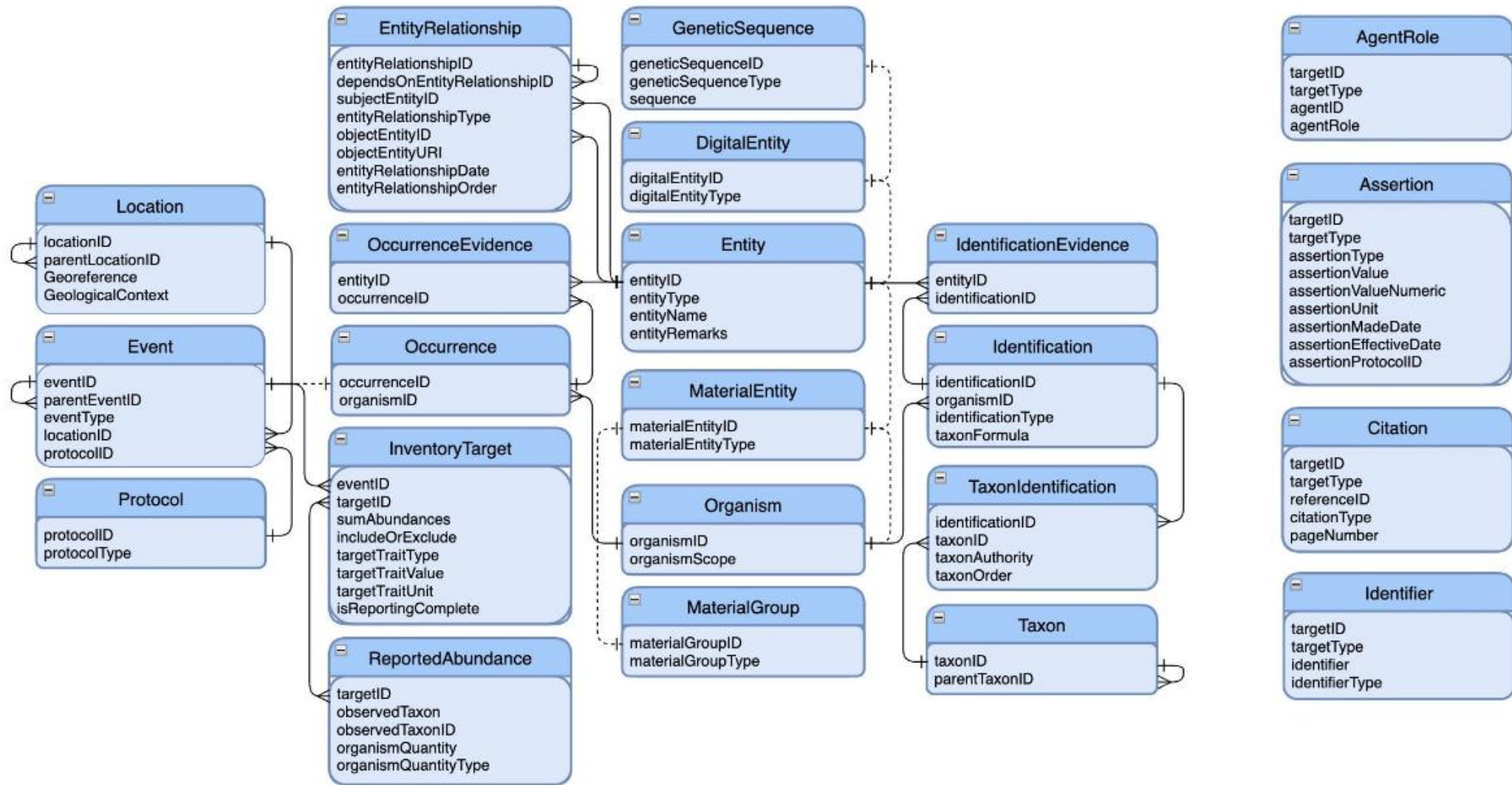
<https://gbif.github.io/dwc-dp/cm/>



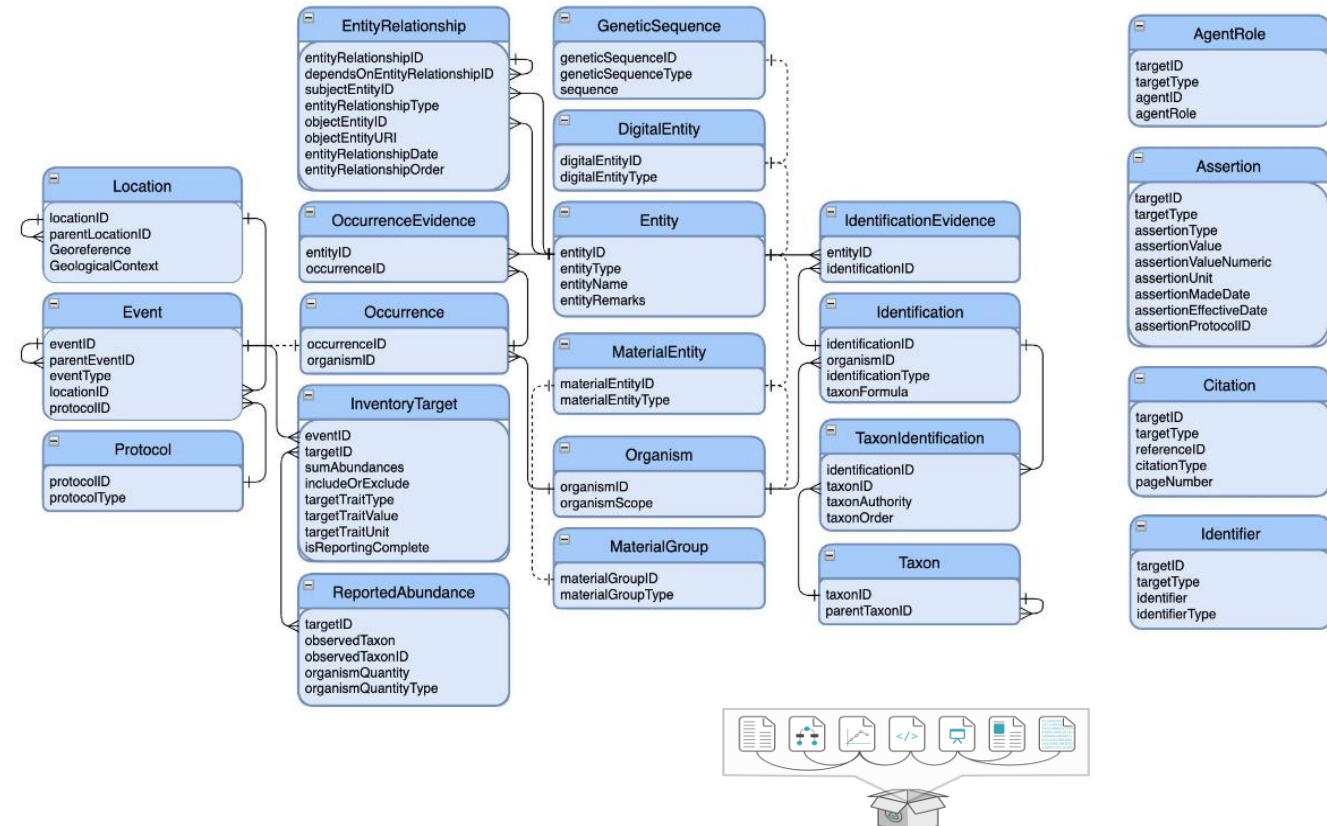
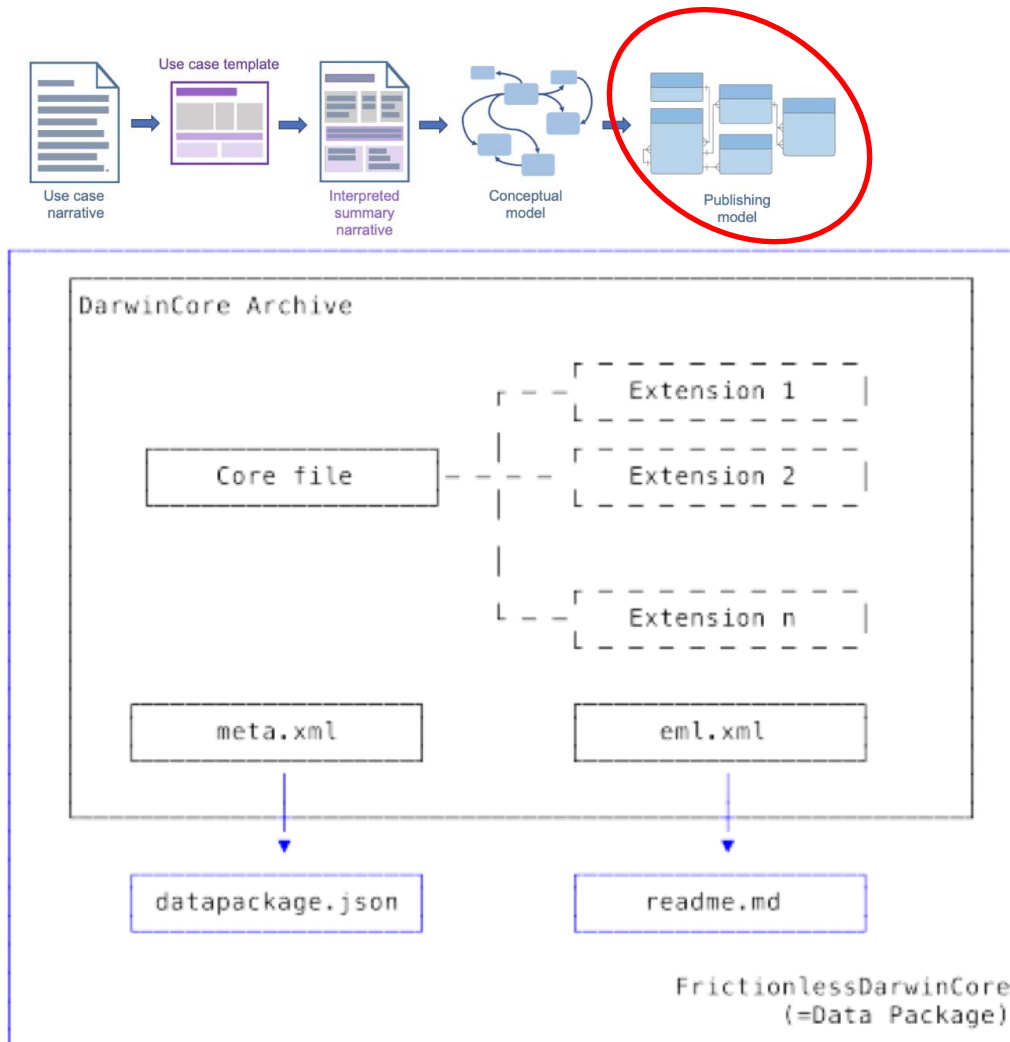
These connect in many-to-many relationships to targets above in context-specific ways



Nuevo modelo conceptual para Darwin Core

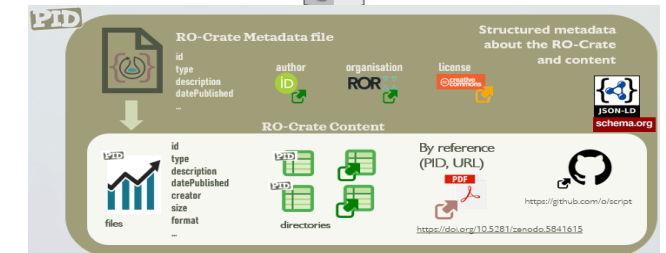


Darwin Core – DP



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

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



Hemos visto...

- Datos y ciencia
- Ciencia abierta y estándares
- Tipos de estándares
- TDWG
- Estándares en GBIF
- Darwin Core
- Plinian Core
- Darwin Core Archive
- Extensiones para colecciones
- GBIF new data model
- Darwin Core – DP

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



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

