



Estandarización y documentación de datos y metadatos

Leonardo Buitrago



August 10, 2026

Agenda

Bienvenida y objetivo

1. Repaso relámpago: Infraestructura AAA + GBIF
2. Captura, procesamiento y calidad de datos
3. ¿Qué es Darwin Core (DwC) y cómo implementarlo?
4. Recorrido por la plantilla ICTIO–DwC
5. Check de comprensión #1
6. Campos estratégicos de metadatos
7. Ejercicio práctico: documentar un desembarco
8. Puesta en común - check final y cierre

TALLERES DE CAPACITACIÓN

Taller 1: Estandarización y documentación de datos y metadatos | 10 AGO

Taller 2: Herramientas para la limpieza y validación de datos | 26 AGO

Taller 3: Citación, atribución y reconocimiento | 16 SEP

Objetivo

Que las organizaciones socias y actores estratégicos de la Alianza Aguas Amazónicas, que apoyan la captura y gestión de datos, comprendan por qué y cómo estandarizar y documentar sus datos bajo Darwin Core, como primer paso hacia su integración a la nueva infraestructura de datos de la AAA.

A large school of piranhas swimming in water. The fish are silvery with a golden shimmer, and their bodies are deep, compressed. They are swimming in a coordinated manner, filling the frame. The water is dark and slightly murky, with some light reflecting off the surface.

1. Infraestructura AAA + GBIF

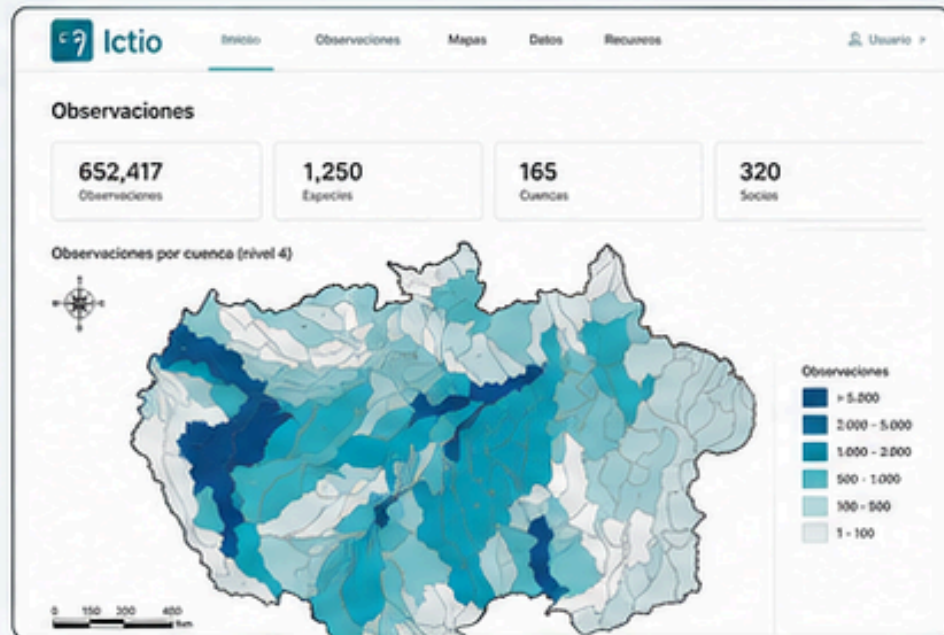
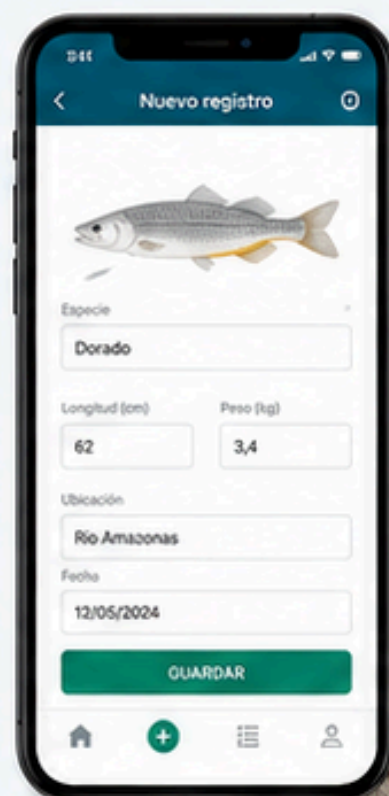


DEL APP AL SISTEMA ICTIO

Reconocer el camino recorrido para entender esta nueva etapa



El camino recorrido



Durante varios años, Ictio fue conocido principalmente como una aplicación móvil y una plataforma web para registrar observaciones de peces y pesquerías en la Cuenca Amazónica.

Pero Ictio
llegó a
representar
mucho más

Sistema Ictio



Un conjunto articulado de herramientas, datos, protocolos, estándares, acuerdos y procesos para generar, organizar, validar y usar información sobre peces, pesca y ecosistemas acuáticos amazónicos.



Esta transición reconoce y proyecta el legado previo

La nueva infraestructura no borra el trabajo de WCS, Cornell Lab of Ornithology y los socios que hicieron posible la etapa inicial de Ictio. Lo ordena, lo proyecta y lo conecta con nuevas herramientas de publicación, visualización y gobernanza de datos.



PUBLICACIÓN

Datos y contenidos
abiertos y reutilizables



VISUALIZACIÓN

Mapas, tableros y
reportes para la acción



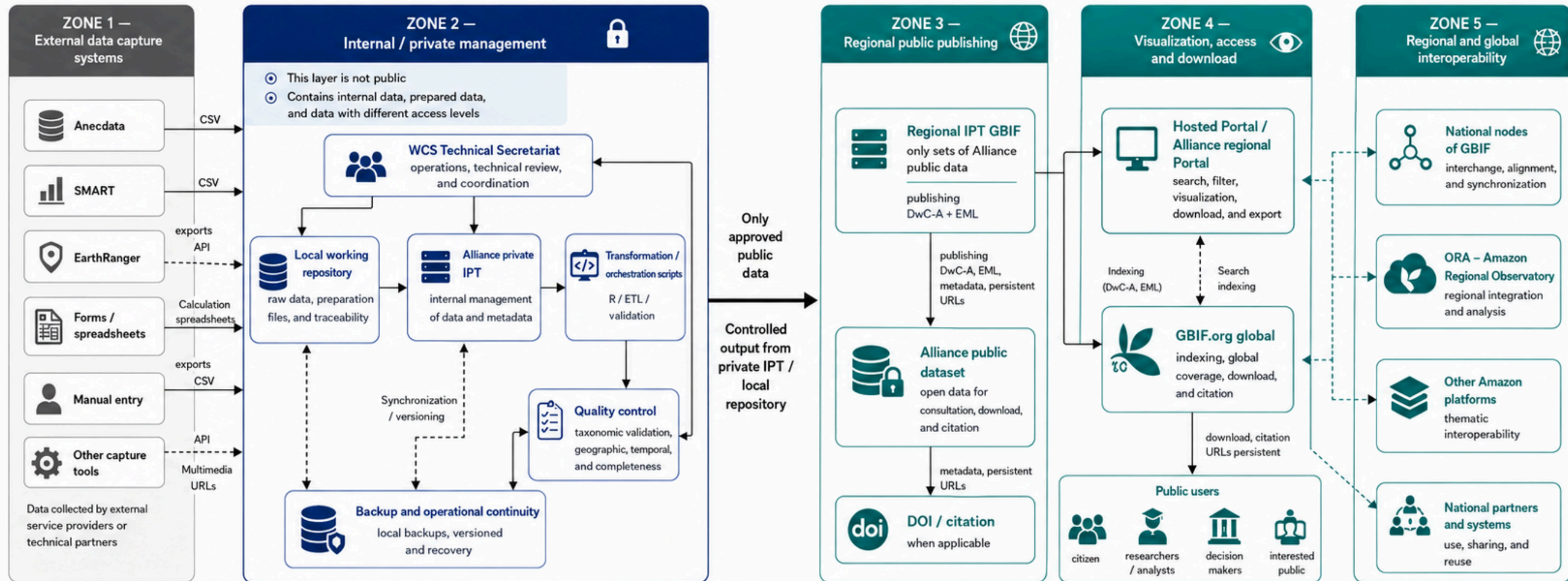
GOBERNANZA DE DATOS

Transparente, participativa
e inclusiva



The Alliance's new technological architecture

Internal management, public publishing, and regional and global interoperability



Features of the new architecture

- ✓ Public access internal management is separated.
- ✓ The Technical Secretariat maintains private IPT management and data preparation.
- ✓ The regional IPT GBIF publishes only the sets of approved public data.
- ✓ GBIF.org and the Hosted Portal enable search, visualization, download, and citation.
- ✓ The architecture improves traceability, quality control, operational continuity, and interoperability.
- ✓ The architecture enables connection with national GBIF nodes, ORA, and other regional platforms.



*Global Biodiversity
Information Facility*

Infraestructura Mundial de
Información sobre Biodiversidad

2.5 mil millones
registros de biodiversidad abiertos

130+ países

participantes como miembros

Fundado en 2001

red intergubernamental libre y abierta

gbif.org

¿Por qué GBIF ?



Datos Abiertos / Principios FAIR

Todos los registros publicados son de acceso libre bajo licencias abiertas (CC0, CC BY). Sin barreras de acceso para investigadores, comunidades o tomadores de decisión.



Reducción de Costos

Infraestructura gratuita: IPT, API, portal alojado y soporte técnico incluidos. No se requiere inversión en servidores ni en desarrollo de plataformas propias.



Interoperabilidad

Conexión con RAISG, ORA, iDigBio, OBIS y cientos de redes globales. Los datos publicados en GBIF son consumibles por cualquier sistema vía API estándar.



Estándares Internacionales

Darwin Core (DwC) y EML para metadatos: los formatos más adoptados en biodiversidad a nivel mundial. Garantiza calidad, trazabilidad y comparabilidad.



Sostenibilidad

Red intergubernamental financiada por gobiernos miembros. No depende de una empresa privada ni de proyectos de corto plazo. Lleva más de 20 años en operación.

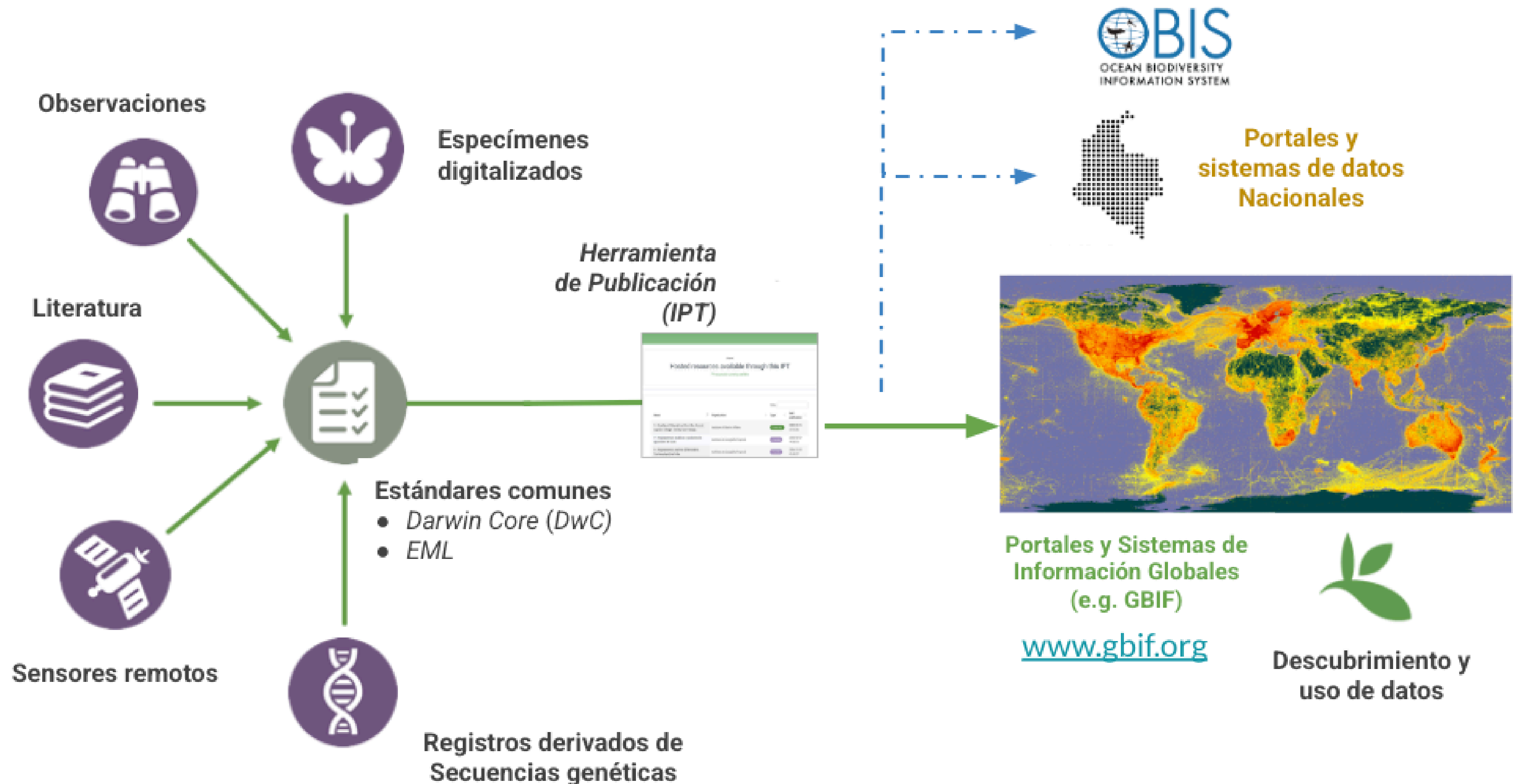


Visibilidad y Ciencia

Más de 10.000 artículos científicos citan datos de GBIF. Publicar aquí aumenta el impacto y la visibilidad de los datos de la Alianza a escala global.

GBIF garantiza que los datos de la Alianza sean abiertos, reutilizables y conectados con la ciencia global

GBIF: UNA VENTANA DE EVIDENCIA SOBRE CUÁNDO Y DÓNDE HAN VIVIDO LAS ESPECIES





2. Captura, procesamiento y calidad de datos

ESTRUCTURA

Flujo de trabajo de datos

Calidad de datos en la captura
de datos

Información taxonómica

Información espacial

Información de colecta

Información descriptiva



DESARROLLAR UN FLUJO DE TRABAJO DE PROCESAMIENTO Y CALIDAD DE DATOS

Cada institución debería contar con:

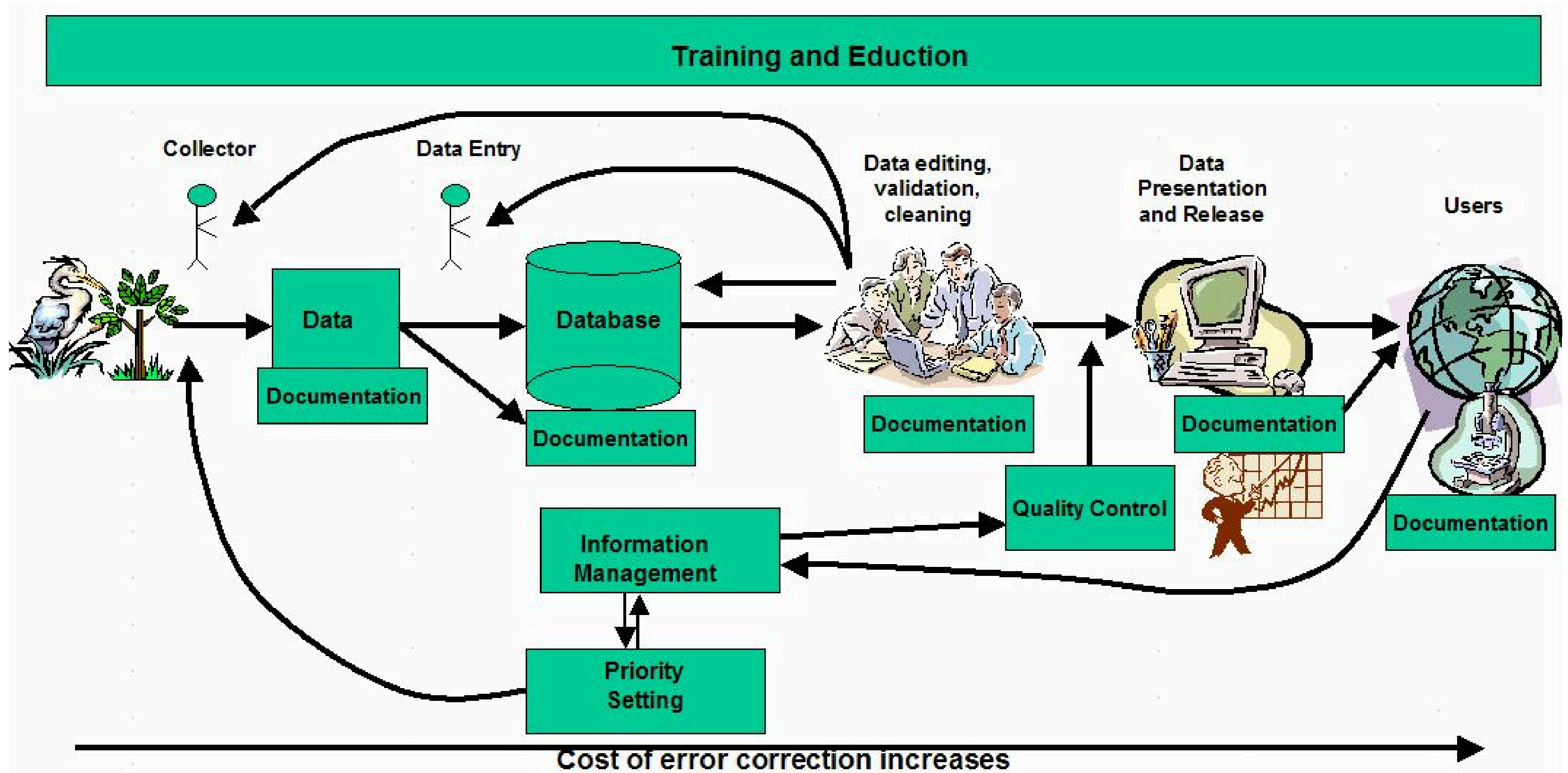
Una visión orientada a la calidad de los datos.

- No «reinventar la rueda»; usar estándares
- Buscar eficiencia (en la recolección de datos y las verificaciones de calidad) y evitar duplicar esfuerzos
- Promover el intercambio (datos, información, herramientas, estándares, etc.)
- Pensar a gran escala
- Atender a los usuarios y sus necesidades
- Invertir en documentación y metadatos

Una política que implemente esta visión.

Una estrategia de implementación para esta política (objetivos precisos a corto, mediano y largo plazo).

FLUJO DE TRABAJO DE PROCESAMIENTO Y CALIDAD DE DATOS



RESPONSABILIDADES DE PROCESAMIENTO Y CALIDAD DE DATOS



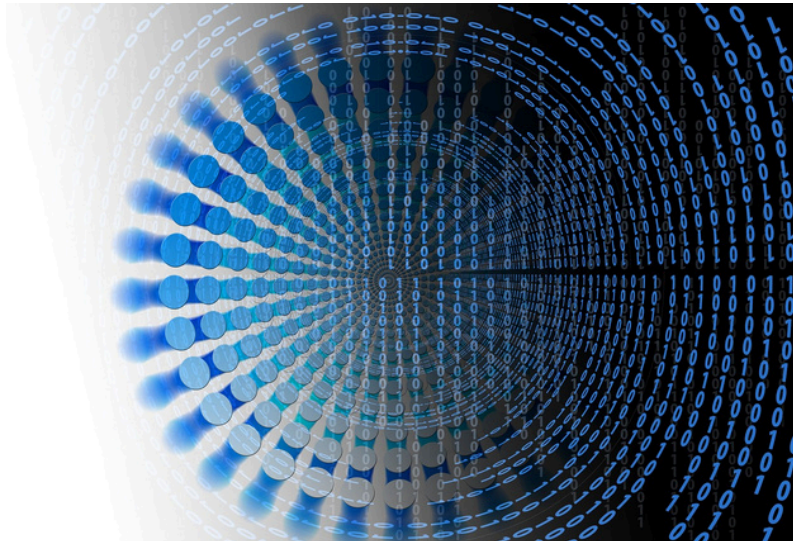
Colector

Proporciona etiquetas y registros legibles, precisos y completos.
Documenta los métodos de colecta
Proporciona observaciones y retroalimentación claras e inequívocas.



Transcriptores

Realiza una entrada de datos precisa.
Performs regular validation tests and corrections.
Realiza copias de seguridad, retención y versionado de archivos de forma periódica.
Atiende la retroalimentación.



Curador

Proporciona metadatos de calidad,
Mantiene un control de calidad y ciclos de retroalimentación regulares.
Reconoce los derechos de propiedad intelectual y las sensibilidades éticas durante las etapas de colecta y publicación.

SEARCH OCCURRENCES 21,439 RESULTS						
IT	MAP	TAXONOMY	METRICS	DOWNLOAD		
ID	Country or area		Coordinates	Month & year	Basis of record	Dataset
naevus, 1758	Cambodia		11.6N, 103.1E	2019 January	Machine observation	WCS Cambodia camera 1
icou (Osbeck, 1765)	Cambodia		11.6N, 103.1E	2019 February	Machine observation	WCS Cambodia camera 1
icou (Osbeck, 1765)	Cambodia		11.6N, 103.1E	2019 February	Machine observation	WCS Cambodia camera 1
ura Linnaeus, 1758	Cambodia		11.6N, 103.1E	2019 March	Machine observation	WCS Cambodia camera 1
ura Linnaeus, 1758	Cambodia		11.6N, 103.1E	2019 March	Machine observation	WCS Cambodia camera 1
ica (J. Geoffroy Saint-Hilaire, 1...	Cambodia		11.6N, 103.1E	2019 March	Machine observation	WCS Cambodia camera 1
vermaphrodite (Pallas, 1777)	Cambodia		11.6N, 103.1E	2019 April	Machine observation	WCS Cambodia camera 1
naevus, 1758	Cambodia		11.6N, 103.1E	2019 May	Machine observation	WCS Cambodia camera 1
naevus, 1758	Cambodia		11.6N, 103.1E	2019 May	Machine observation	WCS Cambodia camera 1
loropus (Blyth, 1859)	Cambodia		12.2N, 107.0E	2018 January	Machine observation	WCS Cambodia camera 1
loropus (Blyth, 1859)	Cambodia		12.2N, 107.0E	2018 January	Machine observation	WCS Cambodia camera 1
loropus (Blyth, 1859)	Cambodia		12.2N, 107.0E	2018 January	Machine observation	WCS Cambodia camera 1
loropus (Blyth, 1859)	Cambodia		12.2N, 107.1E	2018 January	Machine observation	WCS Cambodia camera 1

Usuario

Determina si los conjuntos de datos son adecuados para su uso.
Informa a las colecciones sobre errores y omisiones en los datos.
Proporciona retroalimentación a las colecciones.

ESTRUCTURA

Flujo de trabajo de datos

**Calidad de datos en la
captura de datos**

Información taxonómica

Información espacial

Información de colecta

Información descriptiva



INFORMACIÓN TAXONÓMICA: VOCABULARIO Y CONCEPTOS

Nombres (científicos, vernáculos, rango, jerarquía, etc.)

Estado (sinónimos, nombres válidos, etc.)

Referencias (autor, fecha y lugar)

Identificación (¿por quién y cuándo?)

Términos de calidad (certeza de identificación, etc.)



INFORMACIÓN TAXONÓMICA: ¡CUIDADO CON LOS NOMBRES!

Nombre científico/vernáculo = **principal punto de entrada en una base de datos o portal de datos**

Riesgo de propagar errores durante todo el proceso de publicación de datos

Taxon			
Term	Interpreted	Original	Remarks
Kingdom	incertae sedis		Taxon match none
Accepted name usage ID	DSS004390000Q14K	DSS004390000Q14K	
Scientific name	incertae sedis	Anser anser domestica	Taxon match none
Taxon ID	DSS0043900189RN5	DSS0043900189RN5	
Rank	Kingdom	form	Altered
Vernacular name	DE:Hausgans ,FR:Oie domestique ,	DE:Hausgans ,FR:Oie domestique ,	

Posibles errores y soluciones:

- Identificaciones incorrectas (✓ solicitar ayuda a un taxónomo)
- Errores tipográficos (✓ limpieza de datos)
- Formato incorrecto (✓ limpieza de datos)

Los errores pueden afectar a los nombres científicos y vernáculos, de todos los rangos taxonómicos.

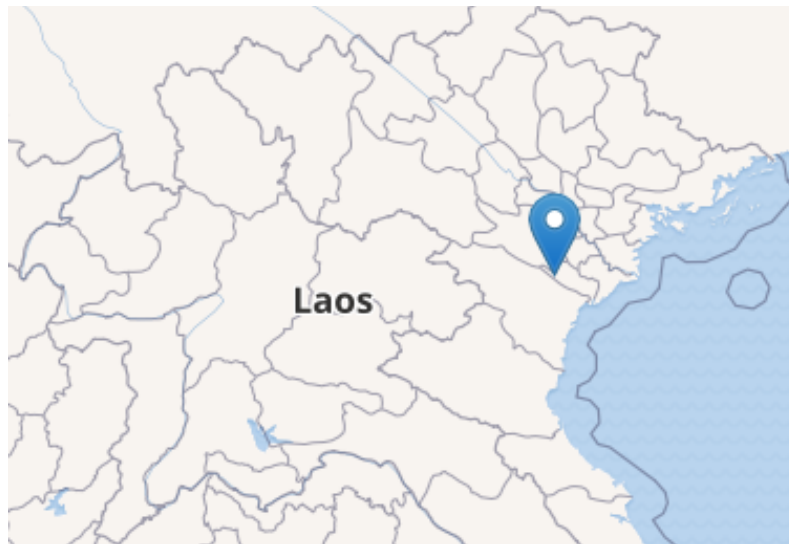
INFORMACIÓN ESPACIAL: VOCABULARIO Y CONCEPTOS



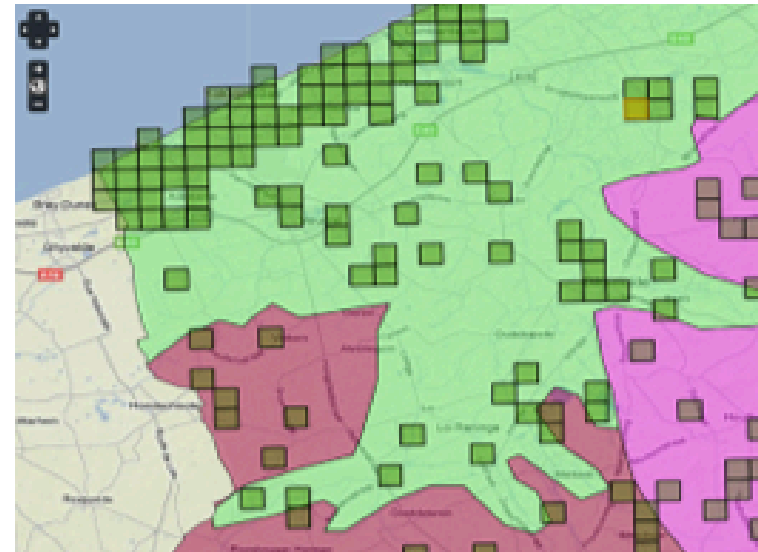
Datos espaciales (textuales o georreferenciados) = información clave para determinar la idoneidad de uso de los datos primarios de biodiversidad para:

- Modelado de distribución de especies
- Selección de áreas a proteger
- Gestión de recursos y del medio ambiente, etc.

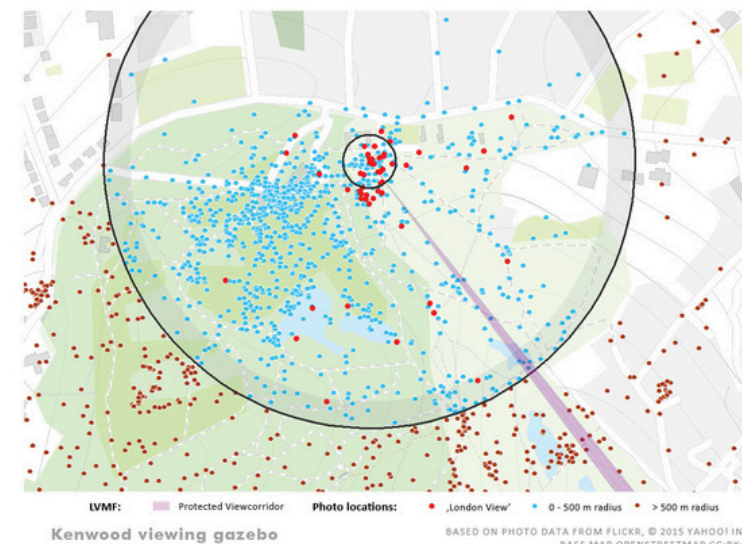
INFORMACIÓN ESPACIAL: ¿DE QUÉ ESTAMOS HABLANDO?



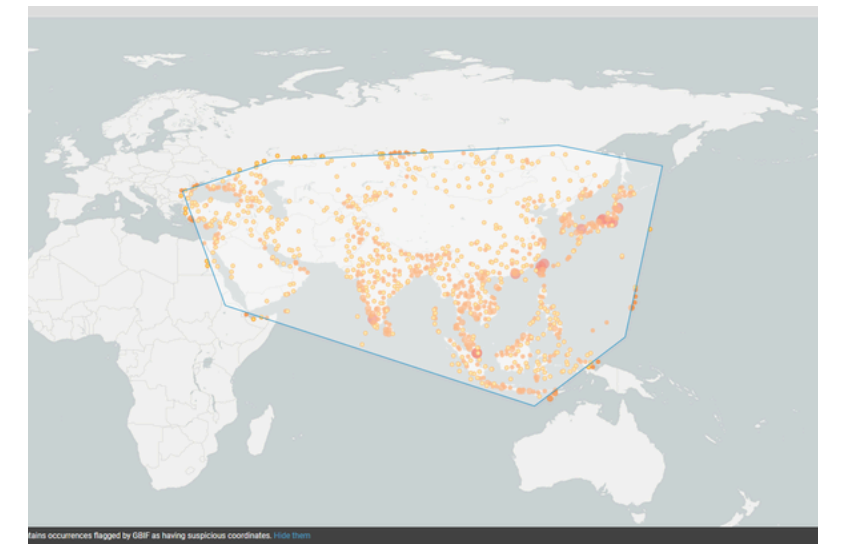
Latitud y longitud



Datos de cuadrícula



Punto + radio



Área/polígono
(+ polilíneas, cuadros
delimitadores...)

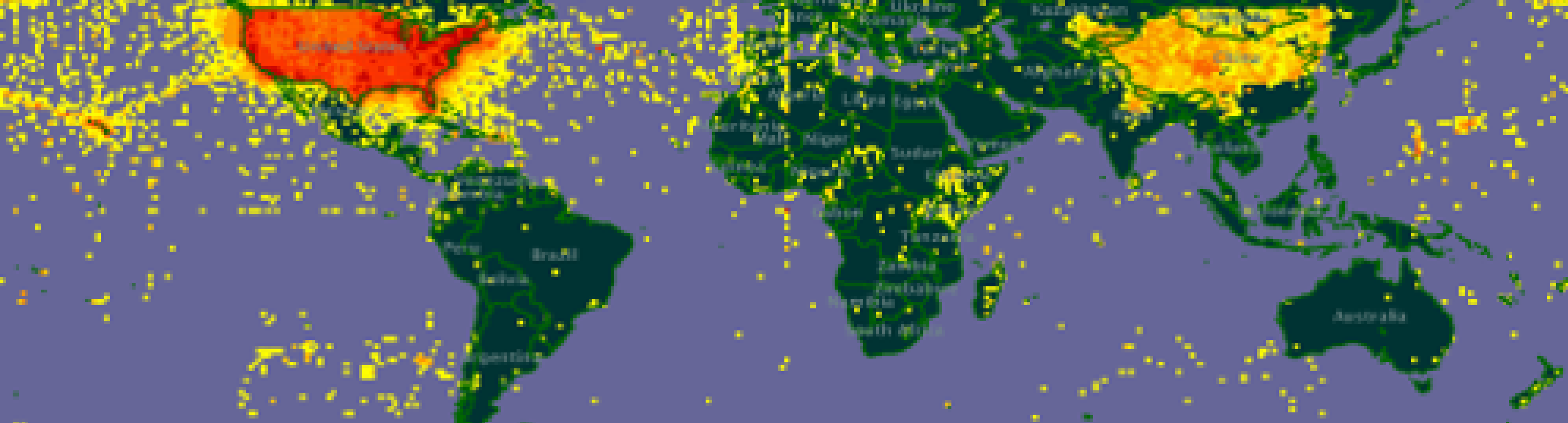
INFORMACIÓN ESPACIAL: ALGUNAS DEFINICIONES MÁS

Coordenadas: un código que documenta una posición en la Tierra, expresada con un SRS (**sistema de referencia espacial**). Suele reportarse como: latitud, longitud y elevación.

Georreferenciación: el proceso de asignar una referencia geográfica a un registro determinado.

Datum geodésico: Un datum geodésico o sistema geodésico es un sistema de coordenadas, y un conjunto de puntos de referencia, que se usa para ubicar lugares en la Tierra (u objetos similares).





Información espacial: errores comunes a evitar

- Inversión de coordenadas
- Valores nulos
- Datum desconocido
- Inadecuado SRS
- Problemas de conversión

INFORMACIÓN DE COLECTA: VOCABULARIO Y CONCEPTOS

Colección de museo: nombre e ID del colector, fecha, hábitat, técnica de captura, etc.

Observaciones: duración de la observación, área, hora del día, actividad, sexo del espécimen observado

Evento de muestreo: métodos de muestreo, tamaño de la cuadrícula, frecuencia, recolección de especímenes de referencia



INFORMACIÓN DE COLECTA: ASPECTOS A TENER EN CUENTA

Exactitud: nombres del/de los colector(es), fecha, etc.

Consistencia: uso de un vocabulario controlado para suelos, hábitats, etc.

Completitud: algunos términos rara vez se completan (período de floración, especies asociadas, etc.), lo que puede dificultar el (re)uso de los datos: intente siempre compartir la mayor cantidad de información posible, si se conoce.



 **GBIF** Darwin Core Vocabulary [\(index\)](#)

Life Stage GBIF Vocabulary

Title Life Stage GBIF Vocabulary
Issued 2015-02-13
URI http://rs.gbif.org/vocabulary/gbif/life_stage
Description A simple vocabulary to represent organism life stages across all kingdoms. Includes quite a few alternative terms
(This is an HTML view of the definition. Use View-Source to see the underlying XML.)

Concepts

zygote A zygote (or zygocyte) describes the first stage of a new unique organism when it consists of just a single cell. The term is also used more loosely to refer to the group of cells formed by the first few cell divisions, although this is properly referred to as a blastomere. A zygote is usually produced by a fertilization event between two haploid cells - an ovum from a female and a sperm cell from a male - which combine to form the single diploid cell. Thus the zygote contains DNA originating from both mother and father and this provides all the genetic information necessary to form a new individual. See also http://rs.gbif.org/vocabulary/gbif/life_stage/zygote

- **zygote** [lang=en, source=]

Code zygote
URI http://rs.gbif.org/vocabulary/gbif/life_stage/zygote
Issued
Alternative Terms

- blastomere [lang=en, source=]

embryo An embryo is a multicellular diploid eukaryote in its earliest stage of development, from the time of first cell division until birth, hatching, or germination.

INFORMACIÓN DESCRIPTIVA: VOCABULARIO Y CONCEPTOS

Calidad variable: datos históricos imposibles de verificar, descripción de datos demasiado costosa en tiempo/dinero, subjetividad (color o la estimación de abundancia, etc.)

A menudo los datos se refieren a todo el rango taxonómico y no al espécimen en particular.

Compleitud: generalmente imposible de lograr en un espécimen determinado.

Consistencia: algunos rasgos pueden ser inconsistentes

FLOWER_COLOUR = Carmine

FLOWER_COLOUR = Crimson



RESUMEN

Planifique todo su **flujo de trabajo de datos**, desde la captura de datos hasta la publicación de datos

La documentación (para reutilizar o compartir datos) es clave en cada etapa del proceso

Las decisiones sobre licencias, estándares, herramientas, etc. deben tomarse a **nivel institucional** y mantenerse **consistentes** para cada conjunto de datos

La calidad de los datos es especialmente importante al capturar y compartir la siguiente información:

- **Información taxonómica** (nombres científicos, taxonomía)
- **Información geográfica** (coordenadas, país, código de país, localidades...)
- **Información de colecta** (cómo y cuándo se observaron y/o recolectaron los especímenes)
- **Información descriptiva** (descripción física de los especímenes y sus atributos)



3. ¿Qué es Darwin Core y cómo implementarlo?

Estándares: acordemos estar de acuerdo



La estandarización no significa que todos vistamos el mismo color y tejido de tela, comamos sándwiches estándar, o vivamos en habitaciones estándar con mobiliario estándar. Se construyen viviendas de infinita variedad de diseño con unos pocos tipos de ladrillos, con madera de medidas estándar, y con tuberías de agua y calefacción y accesorios de dimensiones estándar.

W. Edwards Deming

¿Qué es un estándar?



Convention
Restriction
Rule Requirement
Norm
Specification

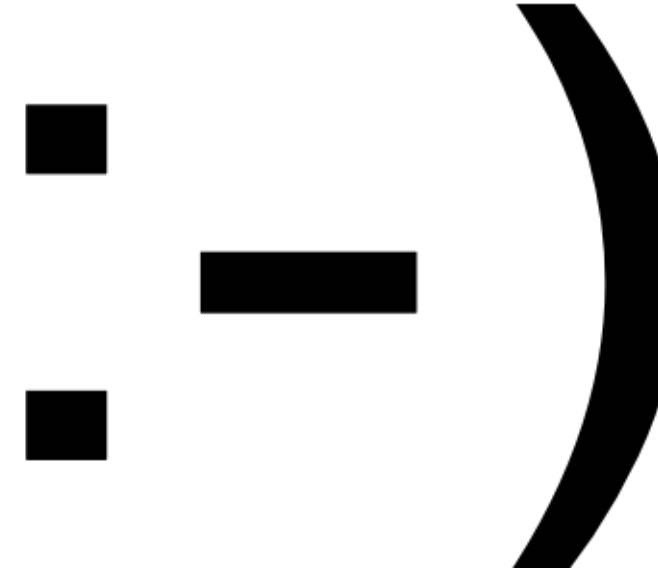
Estándares cotidianos



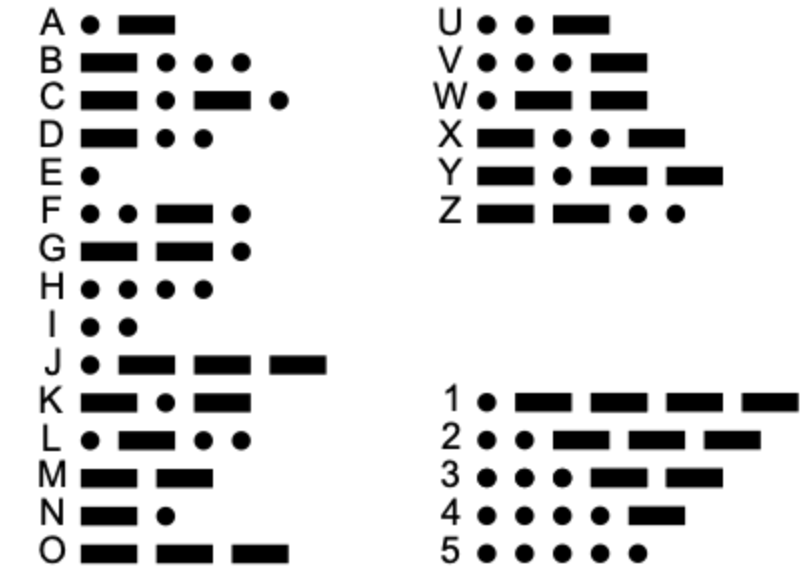
Unidades de medida
(métrico, imperial)



Alfabetos



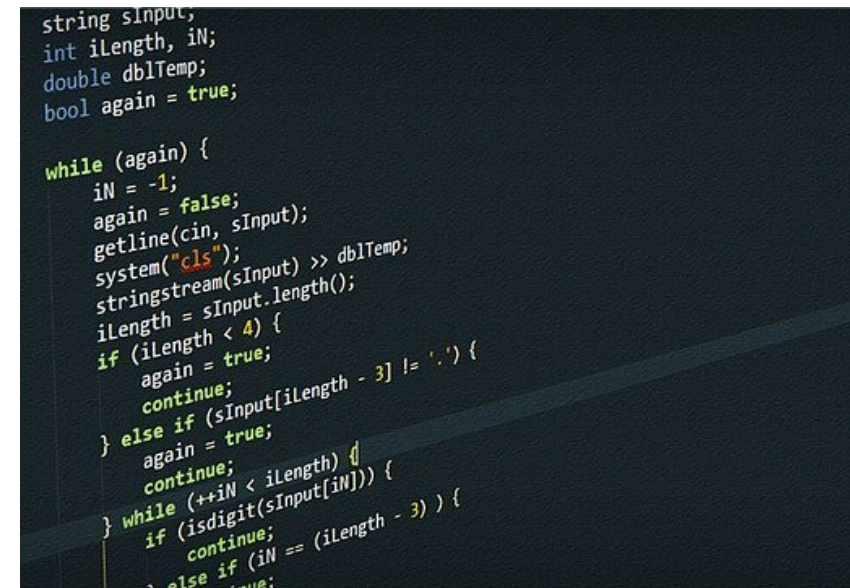
Emojis



Código Morse



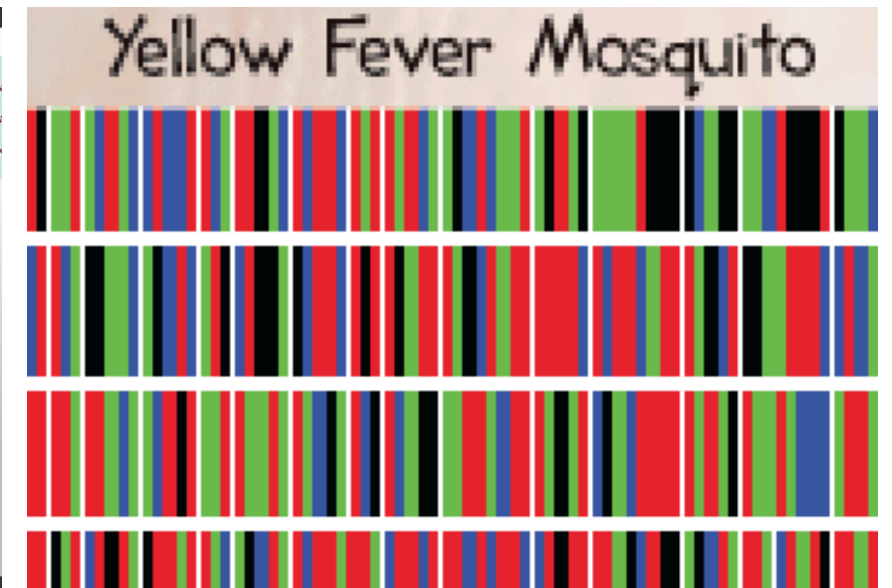
Sistemas numéricos
(indoarábigo; numeración romana)



Idiomas



Direccionamiento postal



Códigos de barras

Un estándar cotidiano – un ejemplo

LATITUD y LONGITUD

- medición - coordenadas geográficas
- formato - grados, minutos, segundos
- sistema numérico - sexagesimal
- números - indoarábigos
- idioma - inglés
- alfabeto - latino
- símbolos - tipografía
- fuente - Roboto



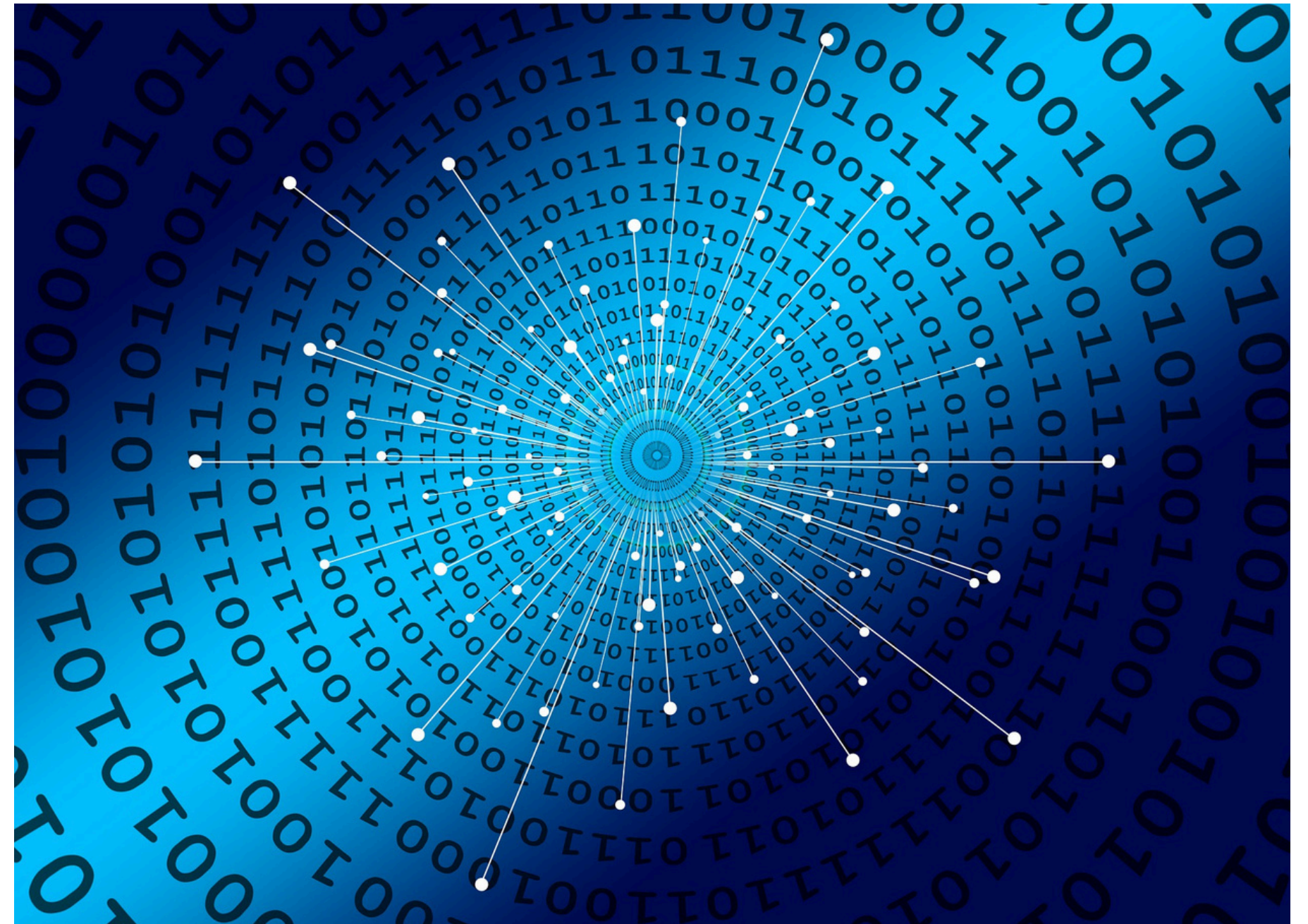
Reglas y restricciones

- Tipo de dato
restricciones sobre la categoría del campo
- Esquema de codificación
restricciones sobre el rango de valores del campo
- Formato
restricciones sobre la representación del dato
- Codificación de caracteres
reglas para interpretar bytes



Estándares para la transferencia de datos

- Esquema de aplicación
Combinaciones específicas de estándares de datos para un propósito particular
- Formato
Restricciones en la estructura del conjunto de datos
- Protocolo de transferencia
Dónde y cómo enviar el contenido



Estándares de información sobre biodiversidad

“Los estándares de datos son las reglas mediante las cuales los datos se describen y registran. Para compartir, intercambiar y comprender los datos, debemos estandarizar tanto el formato como el significado.” (USGS)

Ecological Metadata Language Standard (EML)
Humboldt Ecological Inventory (Humboldt extension)
Global Genome Biodiversity Network (GGBN)
Ocean Data Standards and Best Practices Project (ODSBP)
Darwin Core



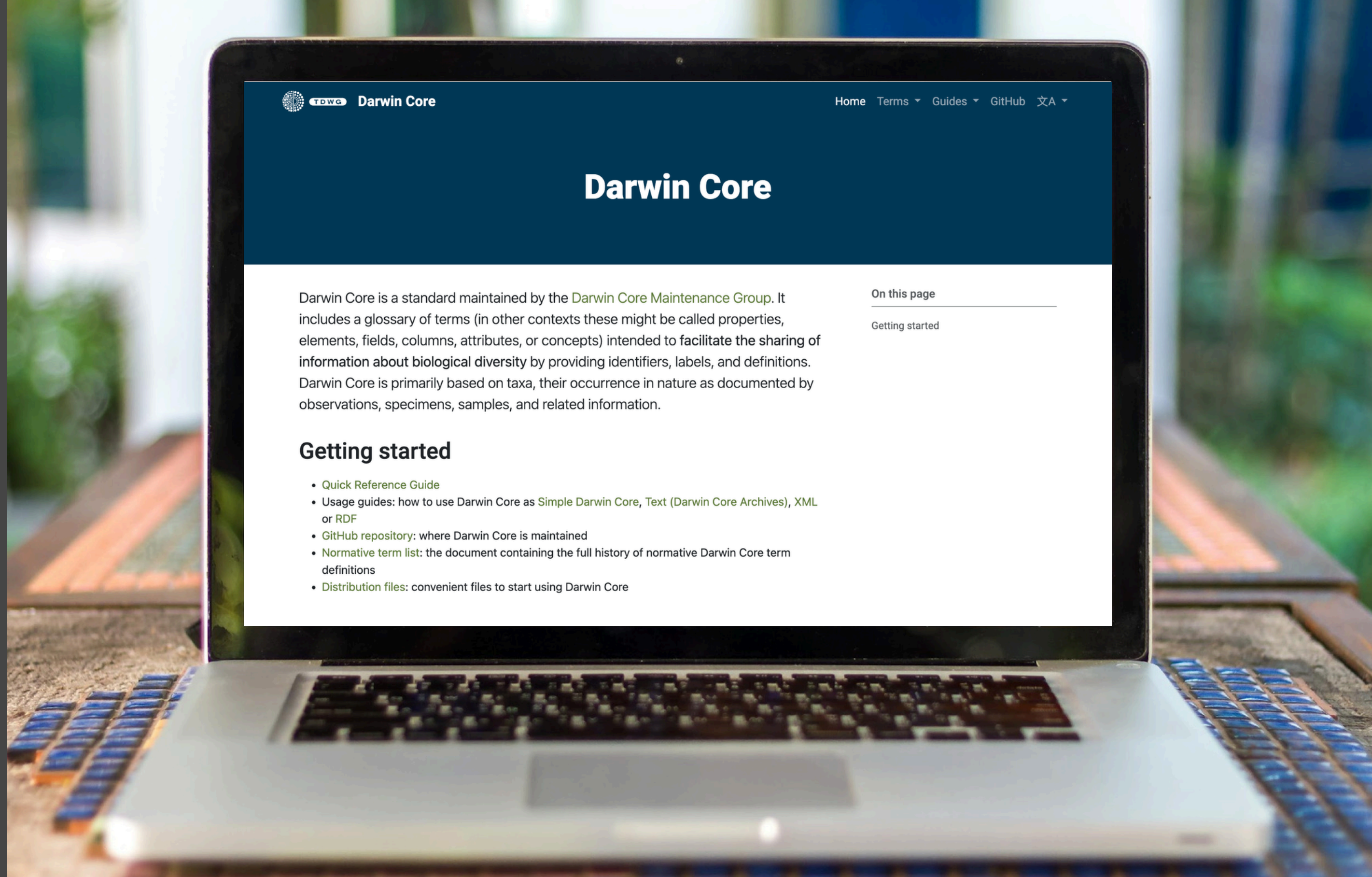
Biodiversity
Information
Standards

TDWG

¿Qué es Darwin Core?

“Una lista de campos y sus definiciones, en relación con los datos de biodiversidad.”

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



Darwin Core Simple

Clases de campos

- Record & Dataset
- Occurrence
- Organism
- Material Entity
- Material Sample
- Event
- Location
- Geological Context
- Identification
- Taxon

Clases auxiliares:

- ResourceRelationship
- MeasurementOrFact



Guía de referencia rápida de Dwc

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

Record-level

This category contains terms that are generic in that they might apply to any type of record in a dataset.

type	modified	language	license	rightsHolder	accessRights	bibliographicCitation
references	feedbackURL	institutionID	collectionID	datasetID	institutionCode	
collectionCode	datasetName	ownerInstitutionCode	basisOfRecord	informationWithheld		
dataGeneralizations	dynamicProperties					

type

Identifier	http://purl.org/dc/elements/1.1/type
Definition	The nature or genre of the resource.
Comments	Must be populated with a value from the DCMI type vocabulary (https://www.dublincore.org/specifications/dublin-core/dcmi-type-vocabulary/2010-10-11/).
Examples	StillImage
	MovingImage
	Sound
	PhysicalObject
	Event
	Text

On this page

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

Términos de DwC: country y countryCode

country	
Identifier	http://rs.tdwg.org/dwc/terms/country
Definition	The name of the country or major administrative unit in which the dcterms:Location occurs.
Comments	Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. Recommended best practice is to leave this field blank if the dcterms:Location spans multiple entities at this administrative level or if the dcterms:Location might be in one or another of multiple possible entities at this level. Multiplicity and uncertainty of the geographic entity can be captured either in the term dwc:higherGeography or in the term dwc:locality, or both.
Examples	Denmark
	Colombia
	España

countryCode	
Identifier	http://rs.tdwg.org/dwc/terms/countryCode
Definition	The standard code for the country in which the dcterms:Location occurs.
Comments	Recommended best practice is to use an ISO 3166-1-alpha-2 country code, or ZZ (for an unknown location or a location unassignable to a single country code), or XZ (for the high seas beyond national jurisdictions).
Examples	AR
	SV
	XZ
	ZZ

On this page

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

Términos de DwC: basisOfRecord

basisOfRecord	
Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use a controlled vocabulary such as the set of local names of the identifiers for classes in Darwin Core.
Examples	MaterialEntity
	PreservedSpecimen
	FossilSpecimen
	LivingSpecimen
	MaterialSample
	Event
	HumanObservation
	MachineObservation
	Taxon
	Occurrence
	MaterialCitation

On this page

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

Términos de DwC: occurrenceID

occurrenceID	
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceID
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.
Comments	Recommended best practice is to use a persistent, globally unique identifier.
Examples	http://arctos.database.museum/guid/MSB:Mamm:233627
	000866d2-c177-4648-a200-ead4007051b9
	urn:catalog:UWBM:Bird:89776

On this page

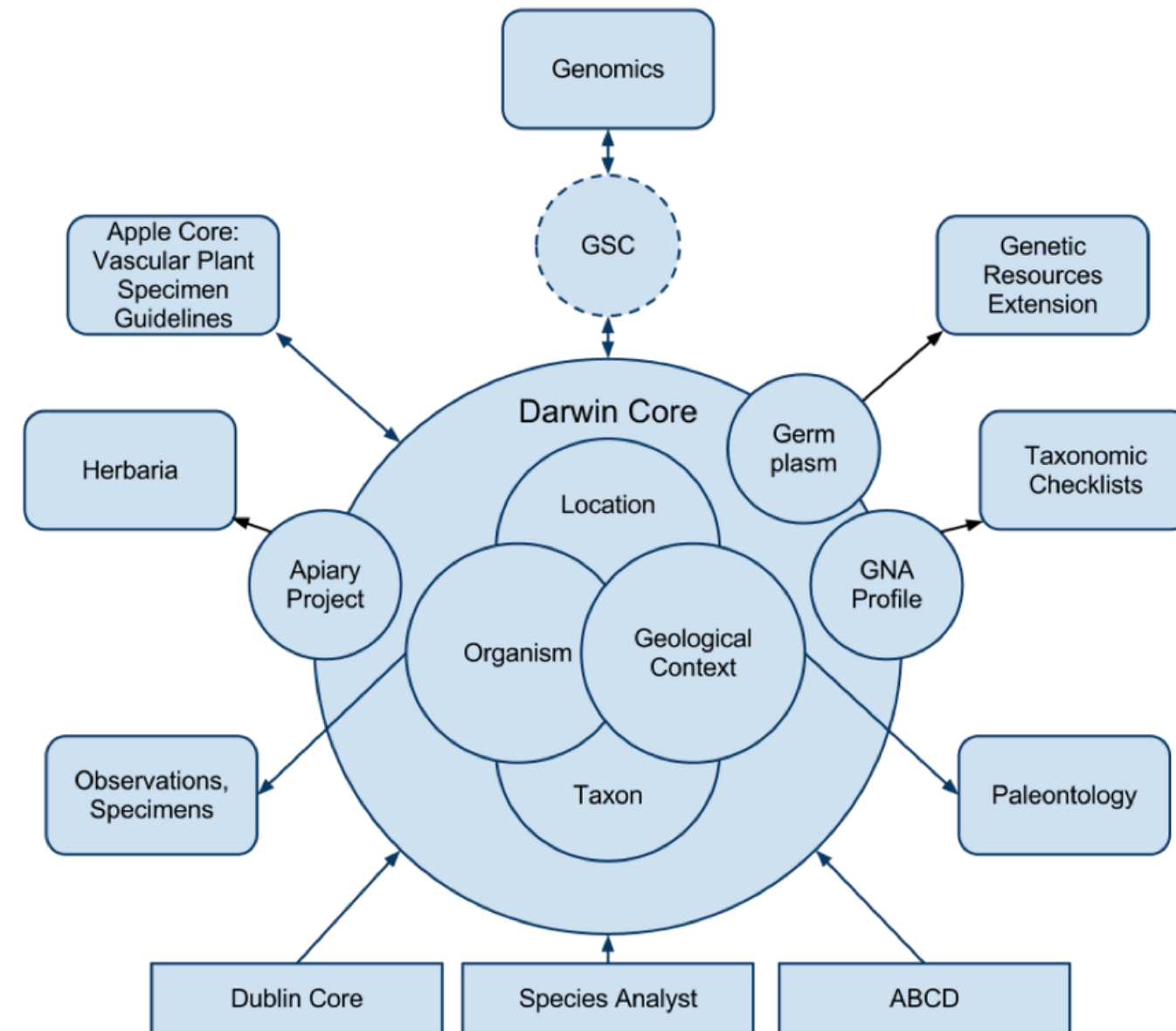
- Record-level
- Occurrence
- Organism
- MaterialEntity
- MaterialSample
- Event
- Location
- GeologicalContext
- Identification
- Taxon
- MeasurementOrFact
- ResourceRelationship
- UseWithIRI
- LivingSpecimen

Extensiones de Darwin Core

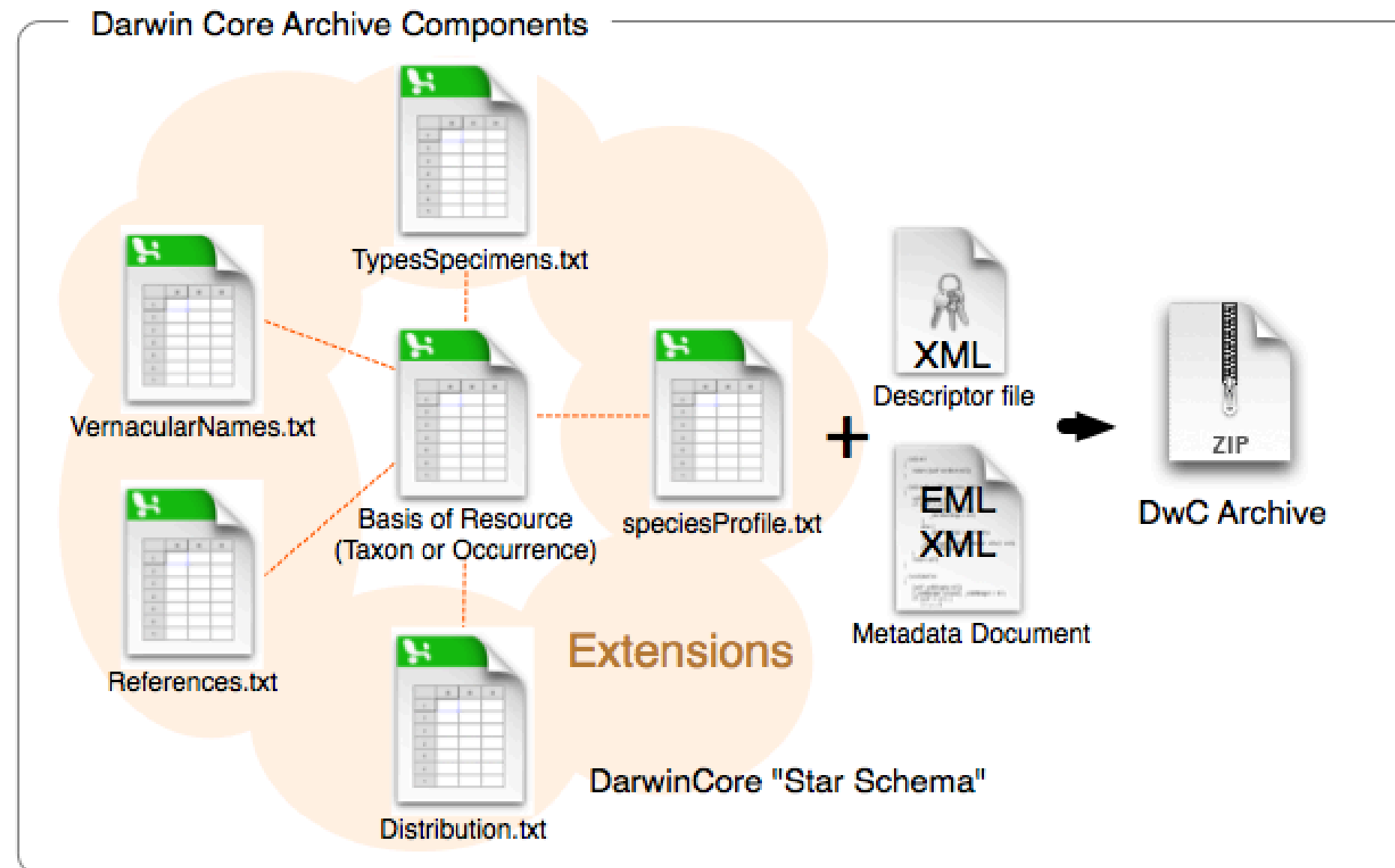
- Simple Multimedia
- Measurements or Facts
- Identification History
- ¡Y muchas más!



Relaciones entre la comunidad y los estándares

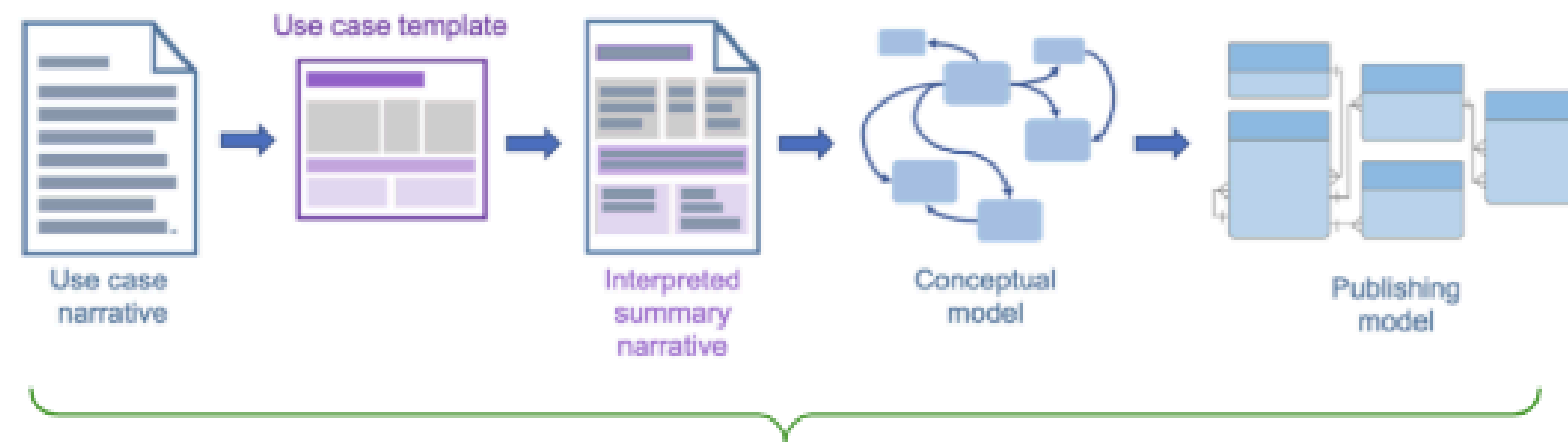


Darwin Core Archives (DwC-A)



Un DwC Archive es una expresión de la guía de texto de Darwin Core. Es un archivo comprimido que contiene un mínimo de tres archivos. Está codificado en UTF-8.

Nuevo modelo de datos de GBIF

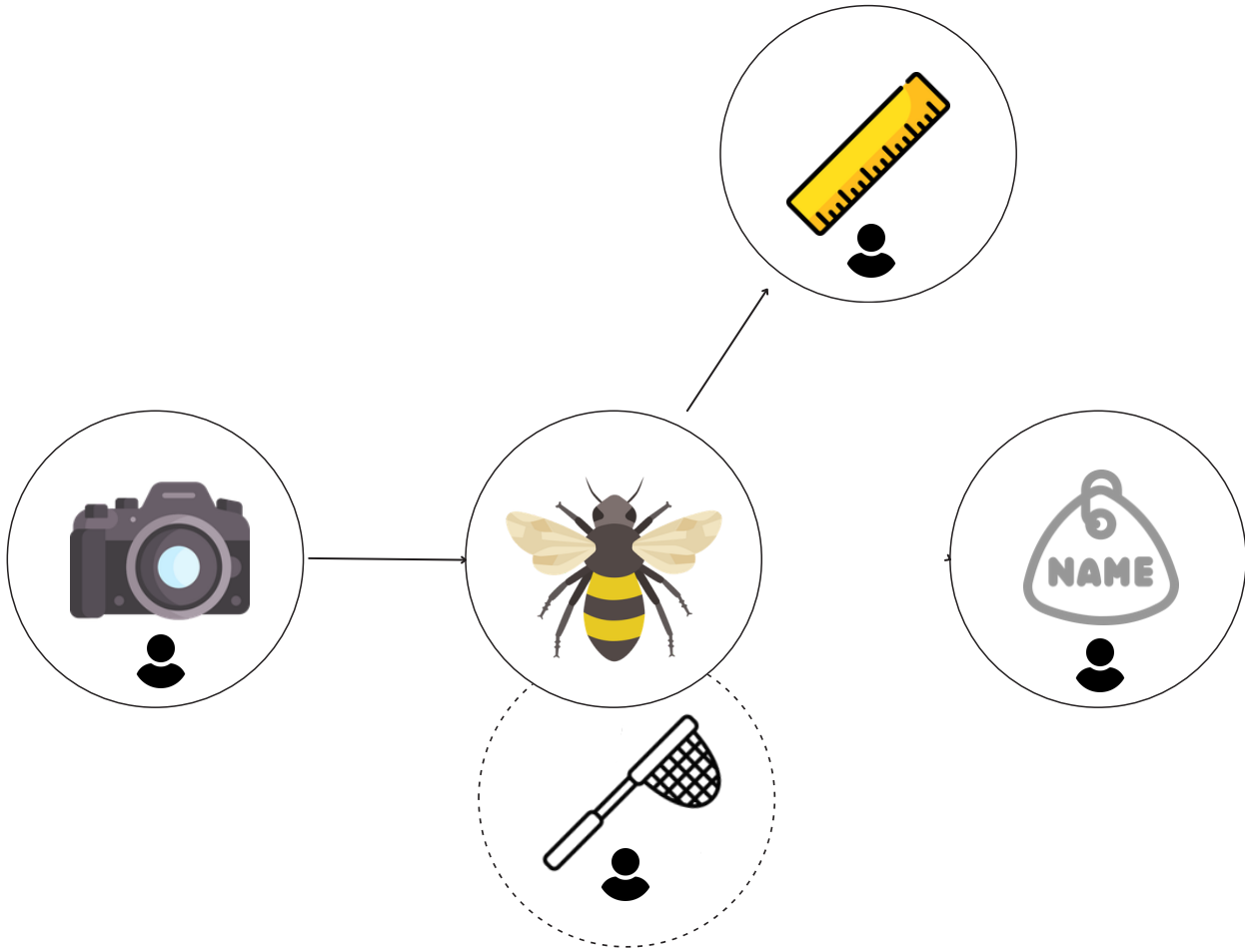


Exploring interactions data on plant pollination	Exploring eDNA data	Exploring camera-trap data	Exploring collection management systems
New data model: Exploring interactions data on plant pollination 📅 20 de septiembre de 2023 🕒 15:00 - 16:00 CEST	New data model: Exploring eDNA data 📅 8 de junio de 2023 🕒 15:00 - 17:00 CEST	Exploring camera-trap data 📅 9 de noviembre de 2022 🕒 15:00 - 17:00 CET	Exploring Collection Management Systems— Webinar 2: Diversifying the data model 📅 7 de julio de 2022 🕒 15:00 - 16:30 CEST

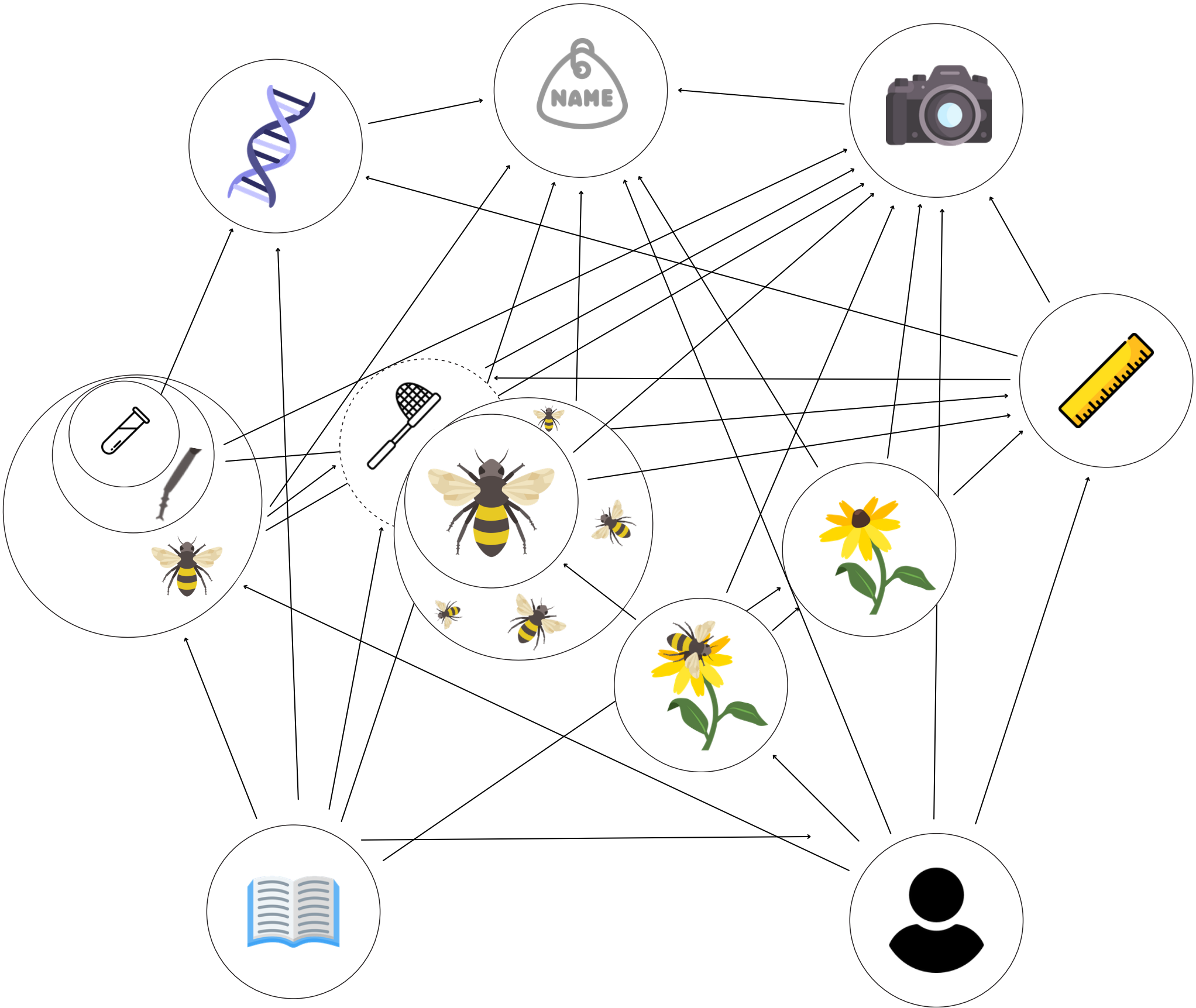


<https://bit.ly/NewDataModel>

Un cuento corto de DwC-A



Una novela épica de DwC-DP

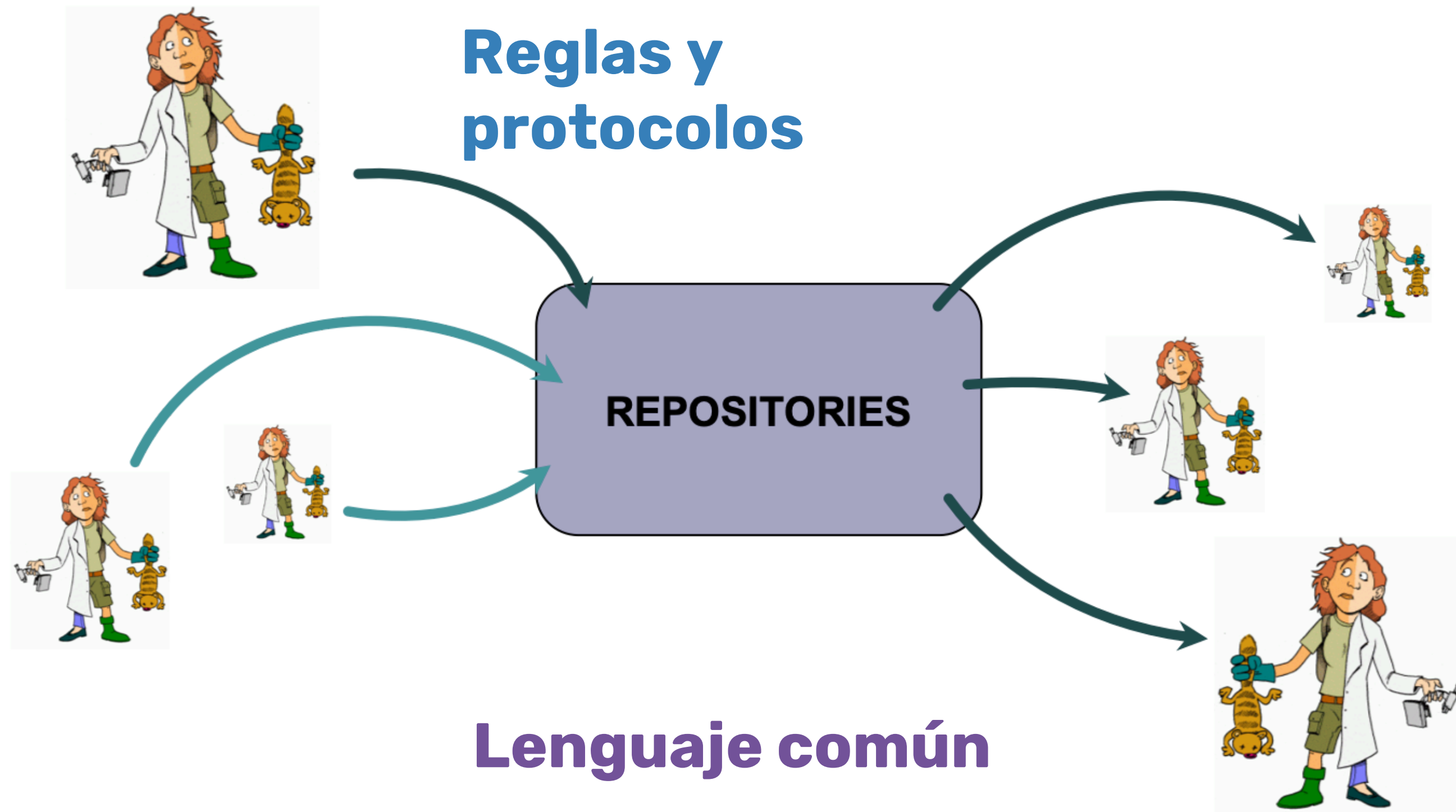


¿Tengo que hacer todo eso?

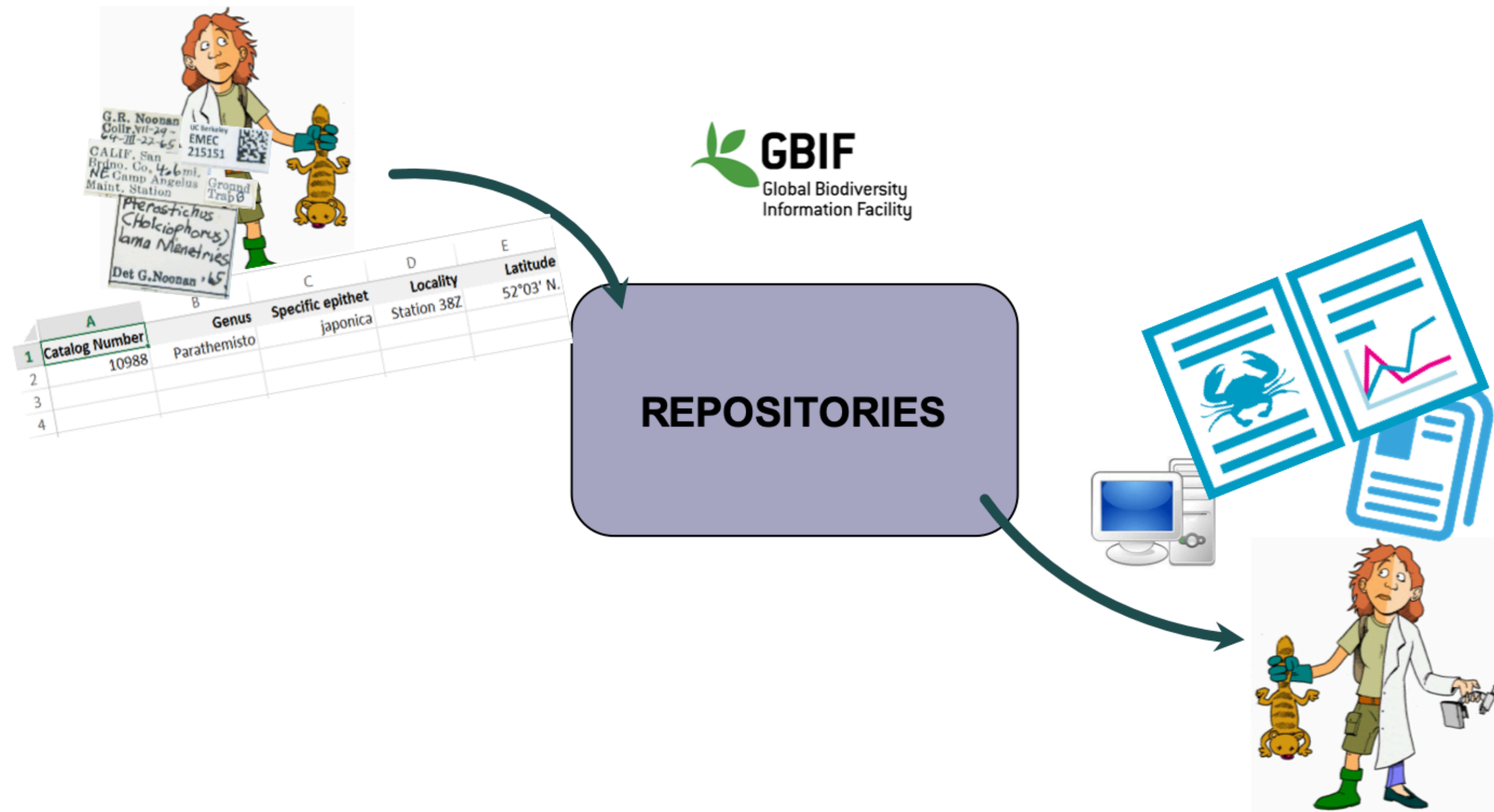
Usen solo lo que necesiten



¿Por qué usar Darwin Core?



¿Por qué usar Darwin Core?



Gracias

<https://training.gbif.org/>



europa.eu

An aerial photograph of a coastal area. A large, irregularly shaped green forested area is the central focus. To the left, there's a road with several vehicles and a cluster of buildings, some with corrugated metal roofs. To the right, more buildings and a road are visible. The foreground shows a dark, possibly paved or wet area. The text '4. Recorrido por la plantilla ICTIO-DwC' is overlaid in white on a dark horizontal band across the middle of the image.

4. Recorrido por la plantilla ICTIO-DwC



<https://tinyurl.com/mrcfm4h>

4. Recorrido por la plantilla ICTIO-DwC

<https://www.menti.com/altjb712dtjm>

5. Check de comprensión #1





<https://tinyurl.com/ybcadhjc>

- 6. Campos estratégicos de Metadatos
- 7. Ejercicio práctico