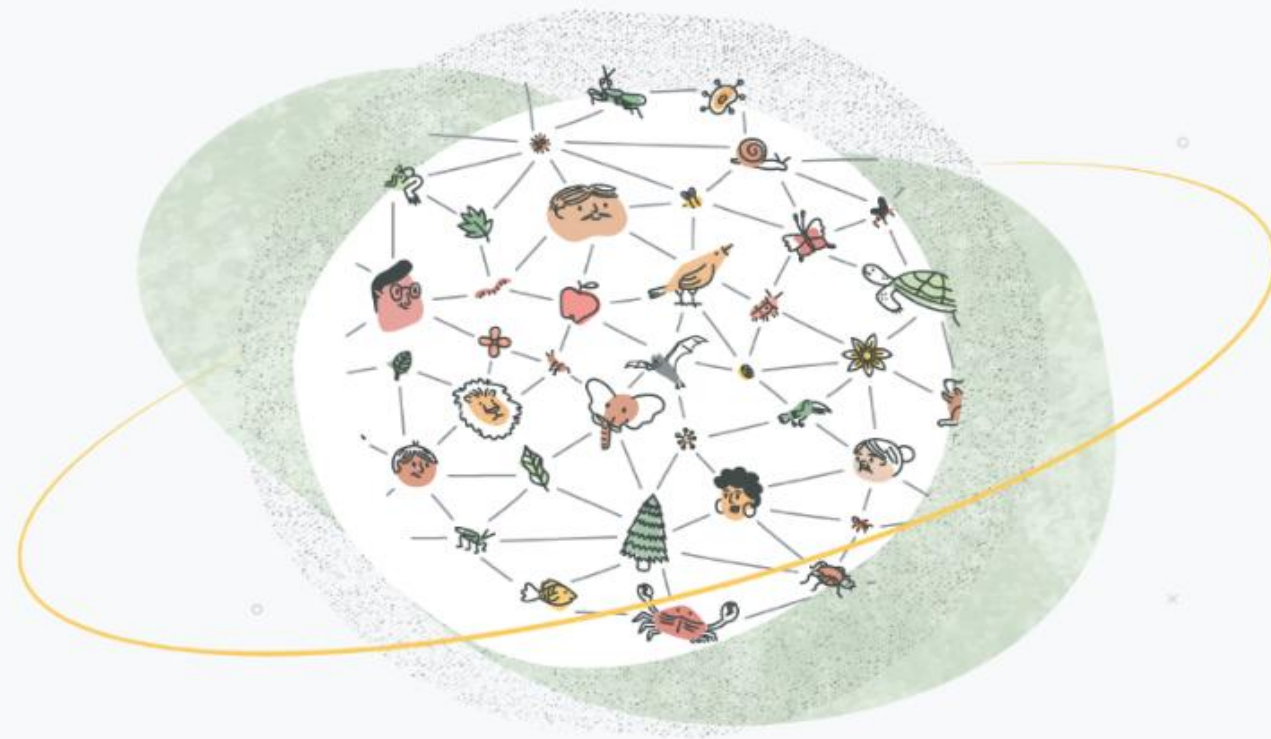


Herramientas para la validación y la estandarización de los datos de biodiversidad



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

Sevilla - Octubre 2025

Katia Cezón katia@gbif.es

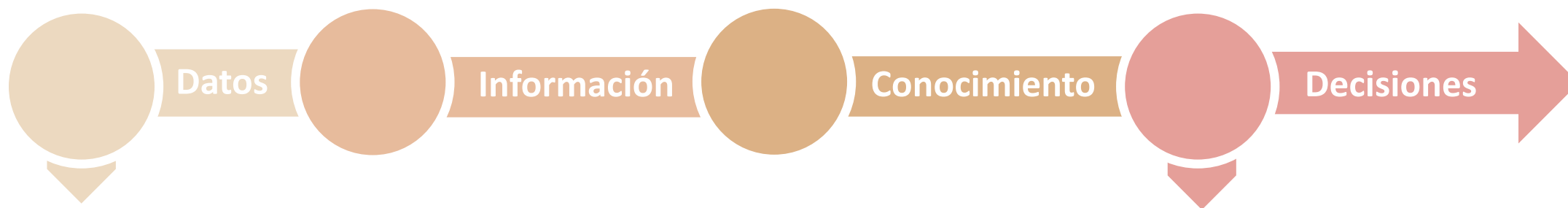


CSIC



Datos primarios de biodiversidad

Fuente de datos como base para el conocimiento

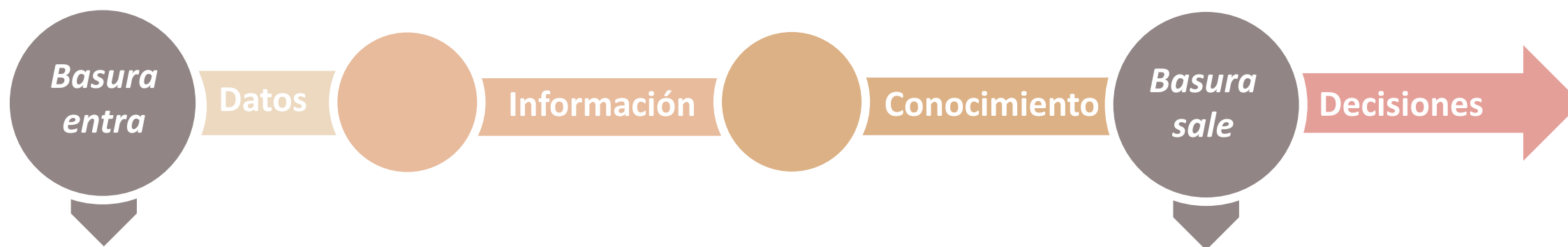


- ✓ Datos digitalizados
- ✓ Localizables
- ✓ Formato accesible
- ? Datos de calidad

Necesitamos un universo de datos lo suficientemente representativo, accesible y de calidad para generar conocimiento y tomar decisiones bien fundamentadas

Datos primarios de biodiversidad

Premisa: Basura entra – Basura sale



Los *problemas de calidad*

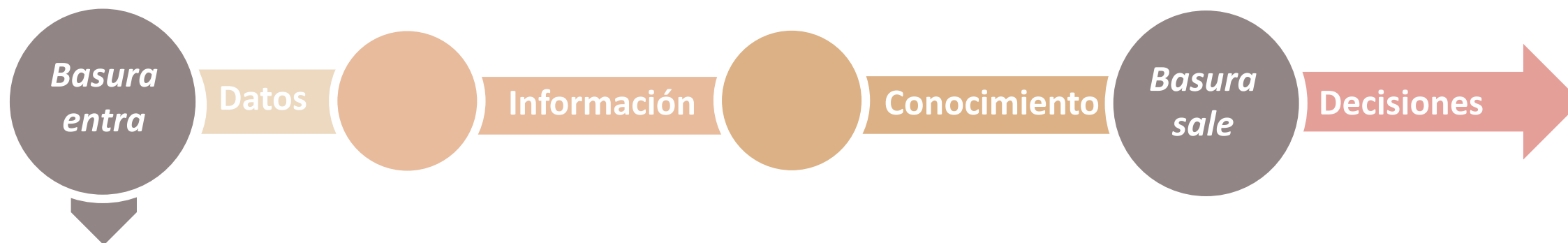
- Dominio taxonómico
- Dominio geoespacial
- Dominio temporal

Impactos

- Investigación científica
- Decisiones de conservación
- Políticas públicas

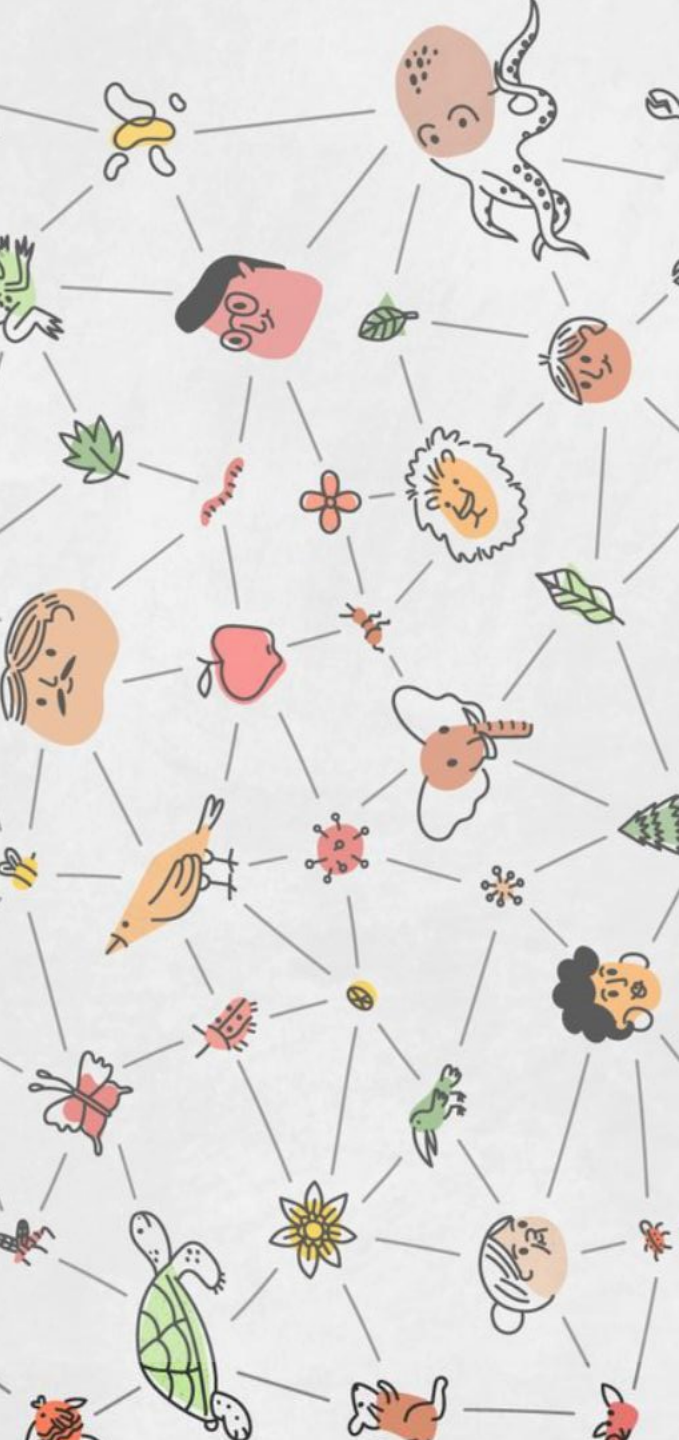
Datos primarios de biodiversidad

Premisa: Basura entra – Basura sale



Los ***problemas de calidad***

- Dominio taxonómico
- Dominio geoespacial
- Dominio temporal



Aspectos a tener en cuenta a la hora de elegir una herramienta

Lo que nos gustaría

Darwin Core Translator

⌘A Texto

📄 Documentos

Mi base de datos

ESPAÑOL FRANCÉS ▼

↔ Darwin Core

CÉS ▼

🎤

0/5000 ✎

Enviar comentarios

Herramientas

Aspectos a tener en cuenta a la hora de elegir una herramienta

No existe la herramienta perfecta

- No existe la herramienta perfecta que haga todo el trabajo.
- Cada herramienta nace para resolver un tipo de problema.
- La elección inteligente está en combinar, no en depender de una sola.

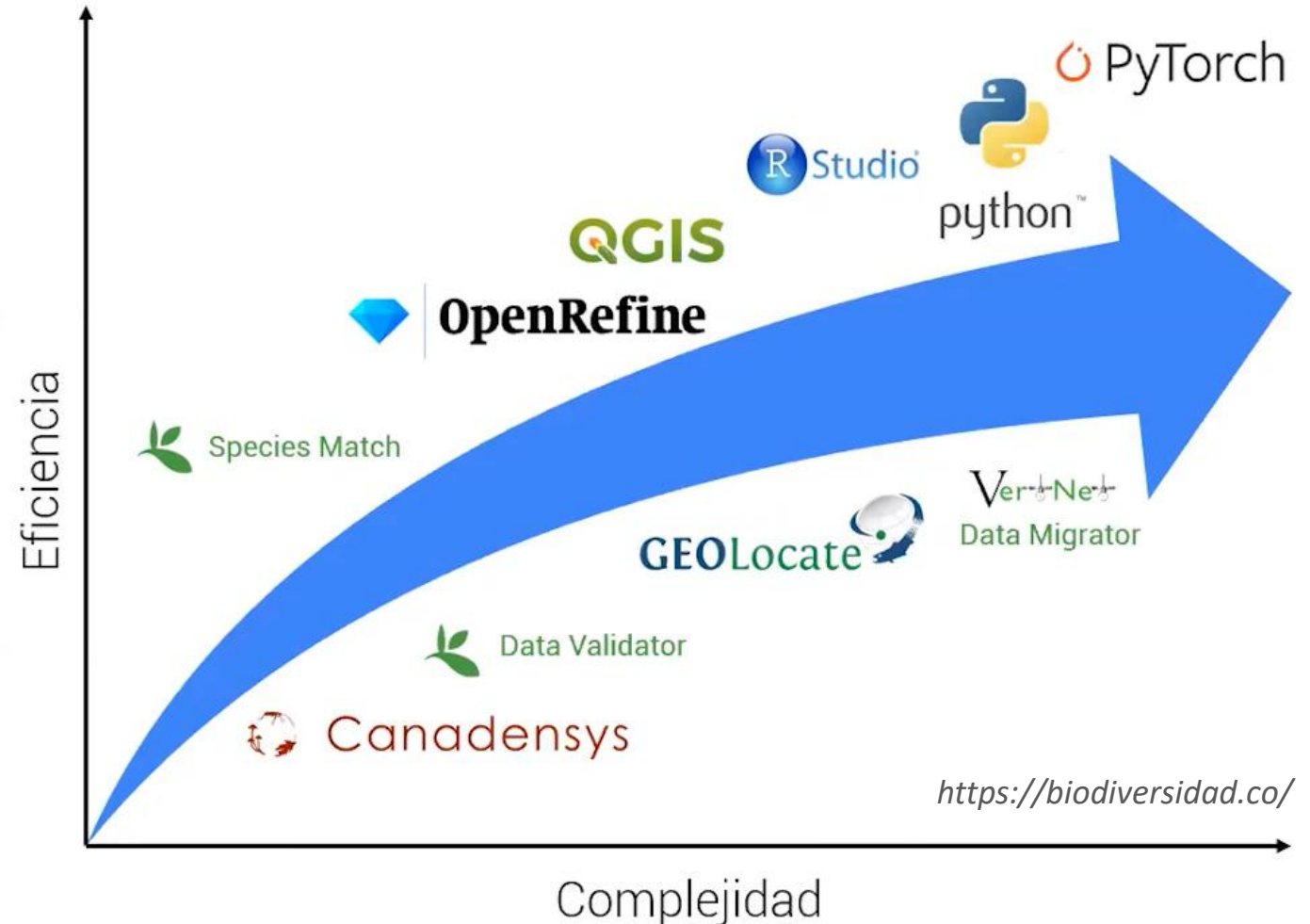


Herramientas

Aspectos a tener en cuenta a la hora de elegir una herramienta

No existe la herramienta perfecta

- No existe la herramienta perfecta que haga todo el trabajo.
- Cada herramienta nace para resolver un tipo de problema.
- La elección inteligente está en combinar, no en depender de una sola.



Herramientas

Aspectos a tener en cuenta a la hora de elegir una herramienta

No existe una receta universal

La elección depende del contexto

- Tipo de datos
- Tamaño del conjunto de datos
- Nivel técnico del equipo
- Infraestructura disponible
- Precio, disponibilidad y licencia.
- Requisitos técnicos: memoria, conexión, instalación.
- Documentación y soporte



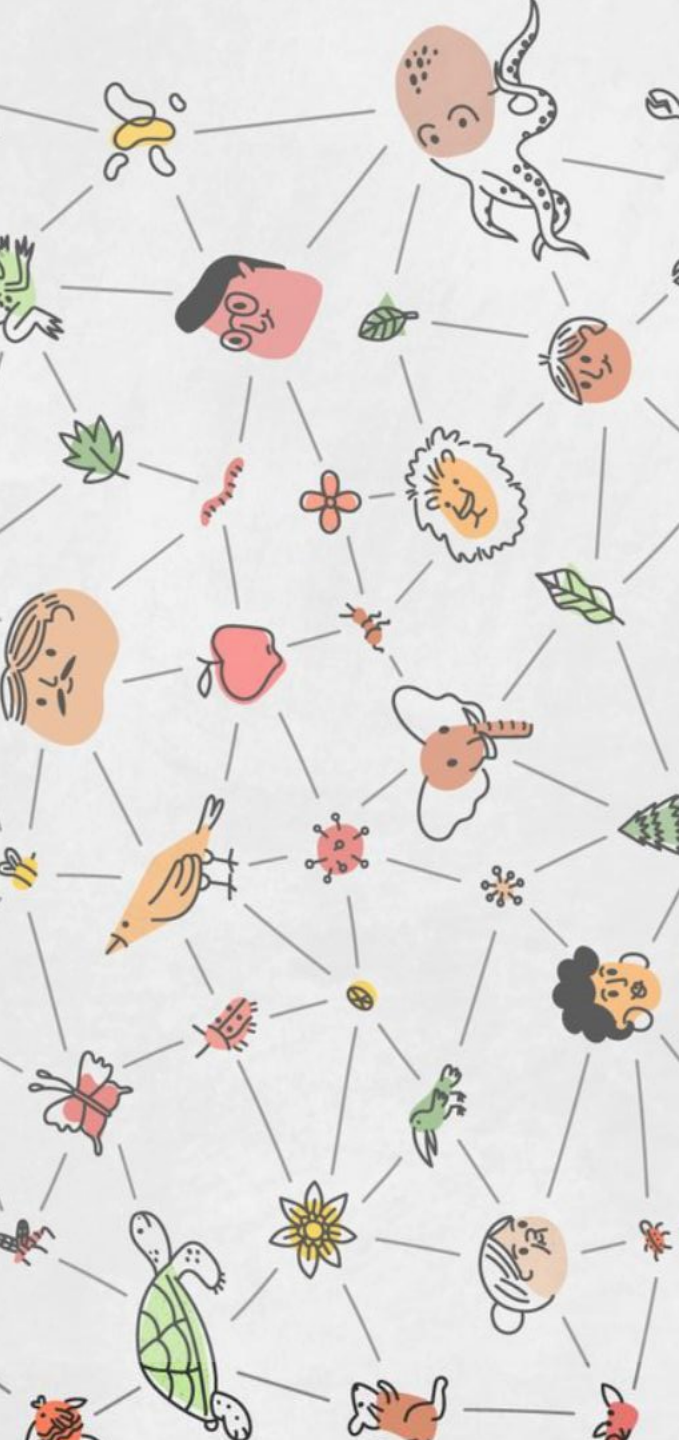
Herramientas

Aspectos a tener en cuenta a la hora de elegir una herramienta

Ninguna herramienta sustituye al criterio humano

Identifique a alguien en su equipo que se sienta cómodo con la manipulación de datos, dado que tendrá que hacer correcciones, migrar desde/hacia diferentes formatos, etc





Retos y herramientas

Dominio temporal

Fechas no estandarizadas

DwC: ISO 8601 YYYY-MM-DD

verbatimEventDate

eventDate

year

month

day


PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS *THE* CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013. II. 27. $27\frac{1}{2}$ -13 2013.158904109
MMXIII-II-XXVII MMXIII ^{LVII}_{CCCLXV} 1330300800
 $((3+3)\times(111+1)-1)\times3/3-1/3^3$ 2013 miss
10/11011/1101 02/27/20/13 $\begin{matrix} 2 & 3 & 1 & 4 \\ 0 & 1 & 2 & 3 & 7 \\ 5 & 6 & 7 & 8 \end{matrix}$ 

Dominio temporal

Fechas no estandarizadas

- 16-mayo-1995
- 16-01-1995
- 16-V-95
- 16/may/1995
- 1995-05-16
- 05/16/1995



DD/MM/YYYY

YYYY-MM-DD

Fecha colecta 16-mayo-1995

→	verbatimEventdate	16-mayo-1995
→	eventDate	1995-05-16
→	year	1995
	month	5
	day	16

Dominio temporal

Herramientas para el tratamiento de las fechas

Específicas



Date parsing



Genéricas



Dominio temporal

Herramientas para el tratamiento de las fechas



Canadensys

Date parsing

Use this tool to parse dates into their component parts. Type or paste dates on separate lines, optionally preceded by your own identifier followed by a tab or a pipe.

Jun 13, 2008

Submit

[Coordinate conversion](#)

[Date parsing](#)

[Tools API](#)

[About](#)

Example input

Jun 13, 2008
15 Jan 2011
2009 IV 02
2 VII 1986

1 | 1999/02/24
2 | 02/17/1921

Dominio geográfico

Estandarización de coordenadas

31TBE65

30T 440 4650

40.3388 -2.0220

30°50'15"N 2°30'10"W 30° 50'15"N 2°30'10"W

30°50'15" N 2 ° 30'1" W

30-50-15 N 2-30-10 W

Geográficas en grados decimales

verbatimCoordinates

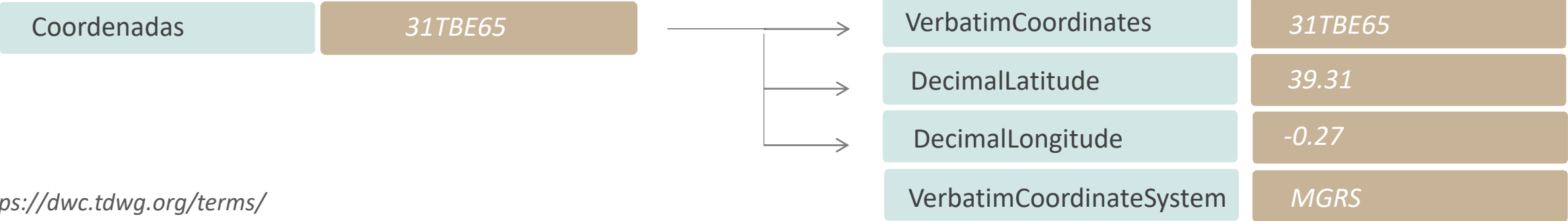
decimalLatitude

decimalLongitude

geodeticDatum

coordinateUncertaintyInMeters

¡Punto “.” como separador decimal!



Herramientas para el tratamiento de los datos geográficos

Conversión de coordenadas a decimales



Earth Point

Dominio geográfico

Estandarización de coordenadas



Convertor de coordenadas

Geotrans

Microsoft Access - Coordenadas a decimal : Base de datos (Access 2007 - 2010)

Herramientas de tabla: Campos, Tabla

Archivos: Archivo, Inicio, Crear, Datos externos, Herramientas de base de datos, Acrobat

Objetos de Access: Tablas (Coordenadas), Consultas (0_Convierte coordenadas), Módulos

CatalogNumber	Coordenadas	coordinateUr	decimalLatit	decimalLongitud
30SXH81		7071	38.06	-0.89
30SXJ6243		707	39.226	-1.117
30SXJ69		7071	39.69	-1.08
30SYH01		7071	38.06	-0.66
31TBF61		7071	40.75	0.22
31TBF60		7071	40.66	0.22
30SXH93		7071	38.24	-0.77
30SXJ6243		707	39.226	-1.117
30TYK05		7071	40.22	-0.59
30SXJ83		7071	39.14	-0.86
30SXX62		7071	39.96	-1.07
30SXH7356		707	38.44	-1.012
30SYH07		7071	38.6	-0.65

Dominio geográfico

Verificación geográfica

Proyecto: Flora vascular del Pirineo

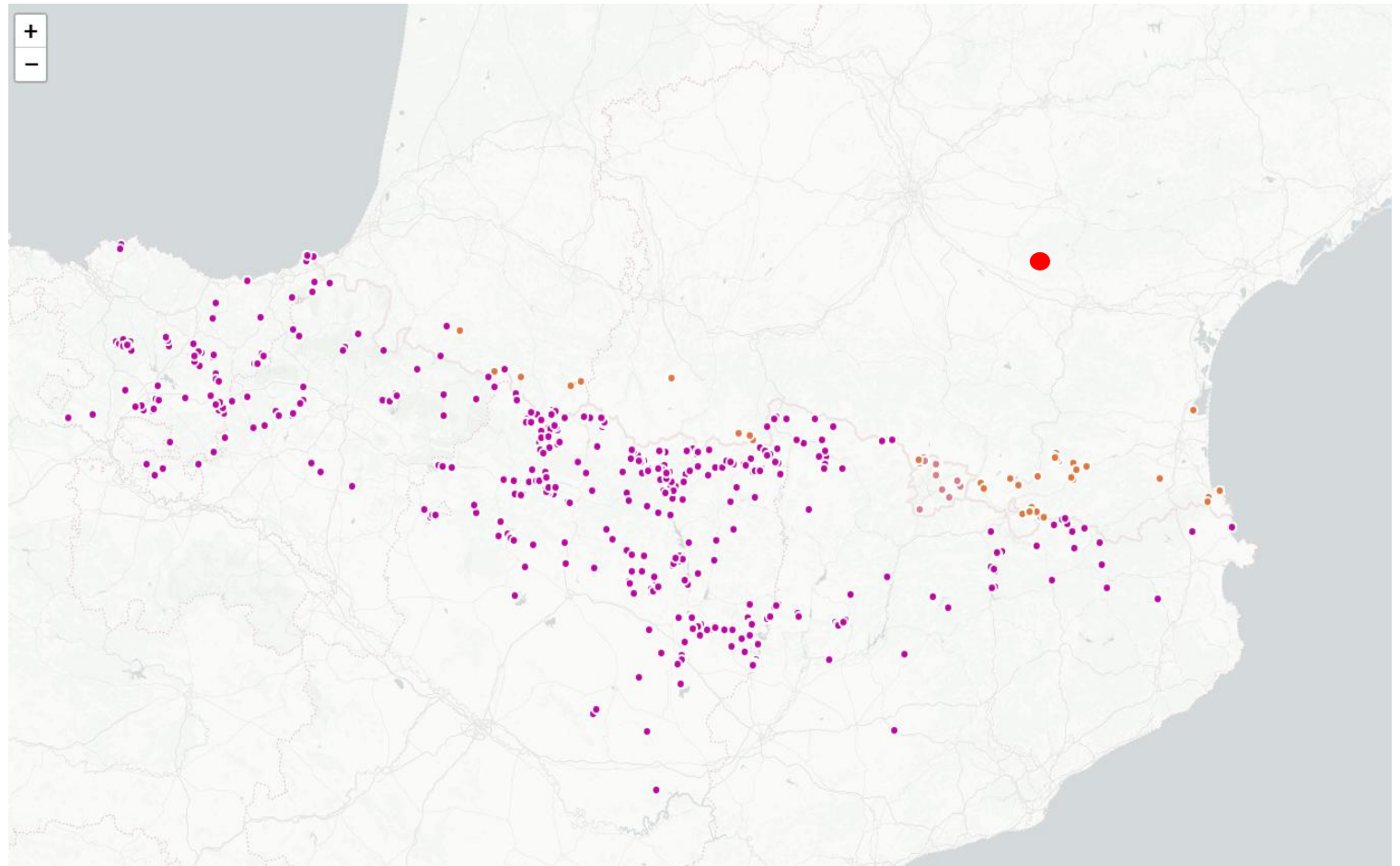
Continent: *Europe*

Country: *Spain*

CountryCode: *SP*

stateProvince: *Álava*

Coordinates: 43.347607 2.097061



Herramientas para el tratamiento de los datos geográficos

Visualización y comprobación

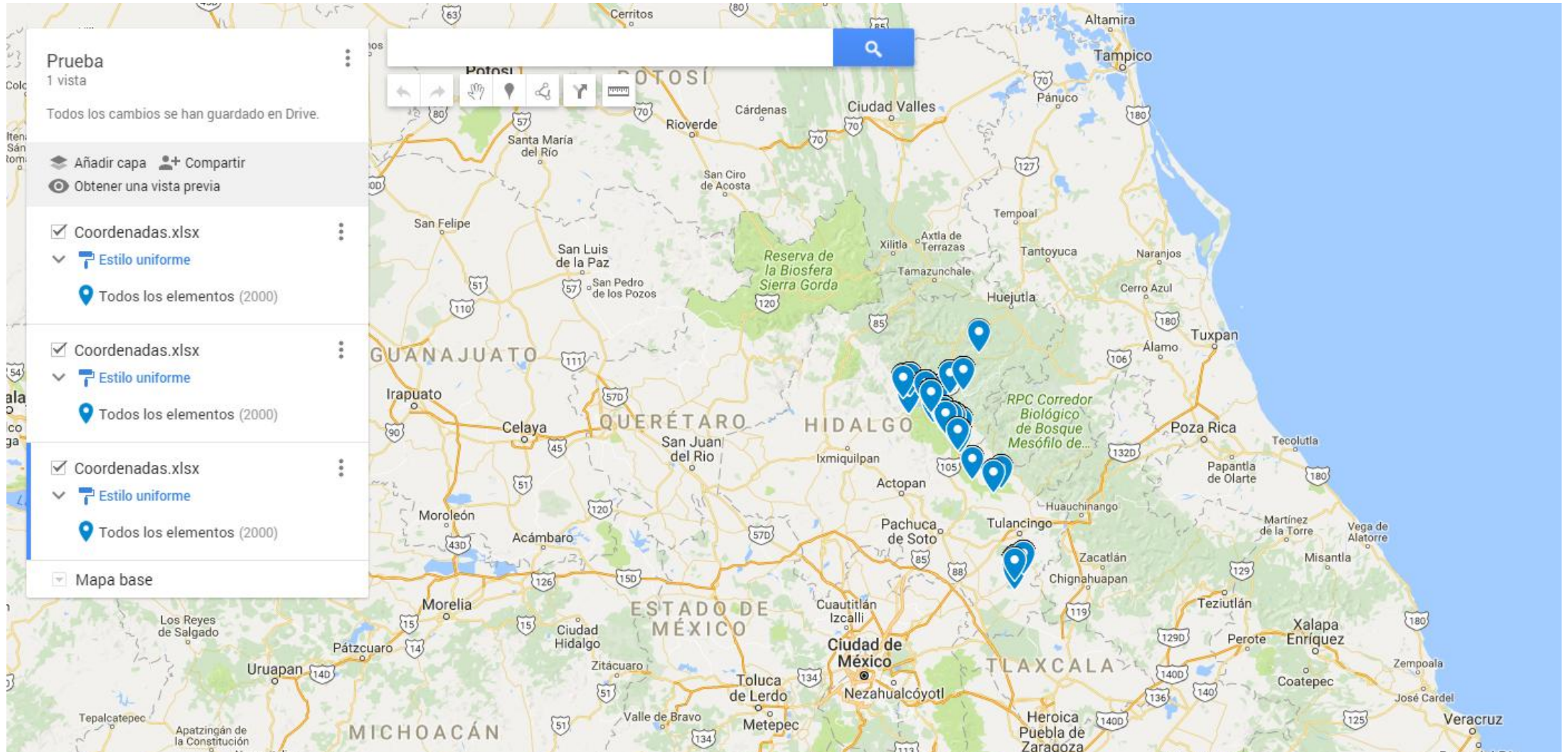


Dominio geográfico

Visualización de datos



Google Maps



Dominio taxonómico

Errores técnicos y de formato

Nombres científicos

scientificName

genus

specificEpithet

infraspecificEpithet

scientificNameAuthorship

taxonRank

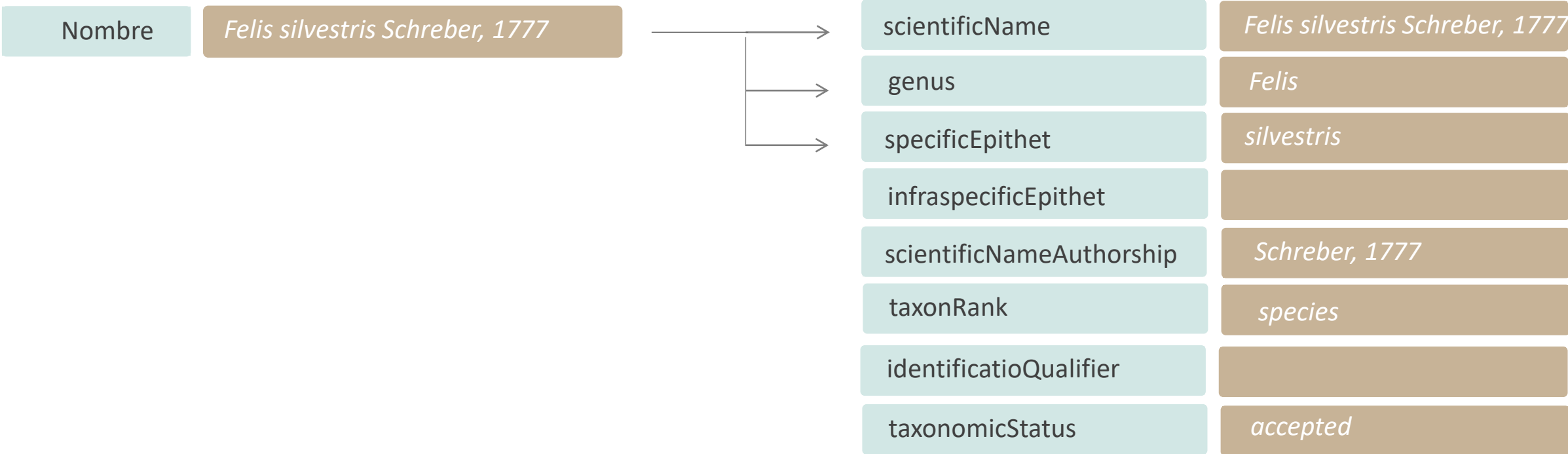
Felis silvestris Schreber, 1777

Felis silvestris Schreber, 1777	Felis Silvestris
Felis sylvestris Schreber, 1777	FELIS SILVESTRIS
Felis silvestris (Schreber, 1777)	Felis sylvestris Linnaeus, 1777
Fesil slvestris	Felis cf. silvestris Schreber
Felis silvestris Schreber 1777	Felis silvestris Schr., 1777
Felis silvestris:Schreber, 1777	Felis Silvestris Schreber,1777
Felis Silvestris SCHREBER, 1777	Felis silvestris (Schreber)
Felis SILVESTRIS Schreber 1777	F. silvestris Schreber, 1777

Dominio taxonómico

Errores técnicos y de formato

Felis silvestris Schreber, 1777



Dominio taxonómico

Errores de consistencia

Felis silvestris Schreber, 1777

¿El registro corresponde realmente a
Felis silvestris? ¿Fue el ejemplar
correctamente identificado?

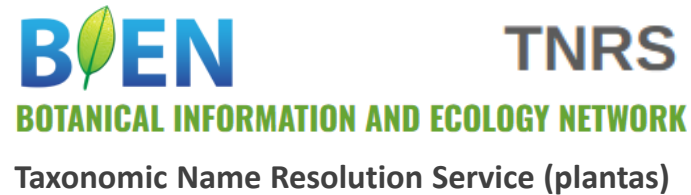


Herramientas para el tratamiento de los nombres científicos

Atomización (Name parser)



Resolución



TNRS



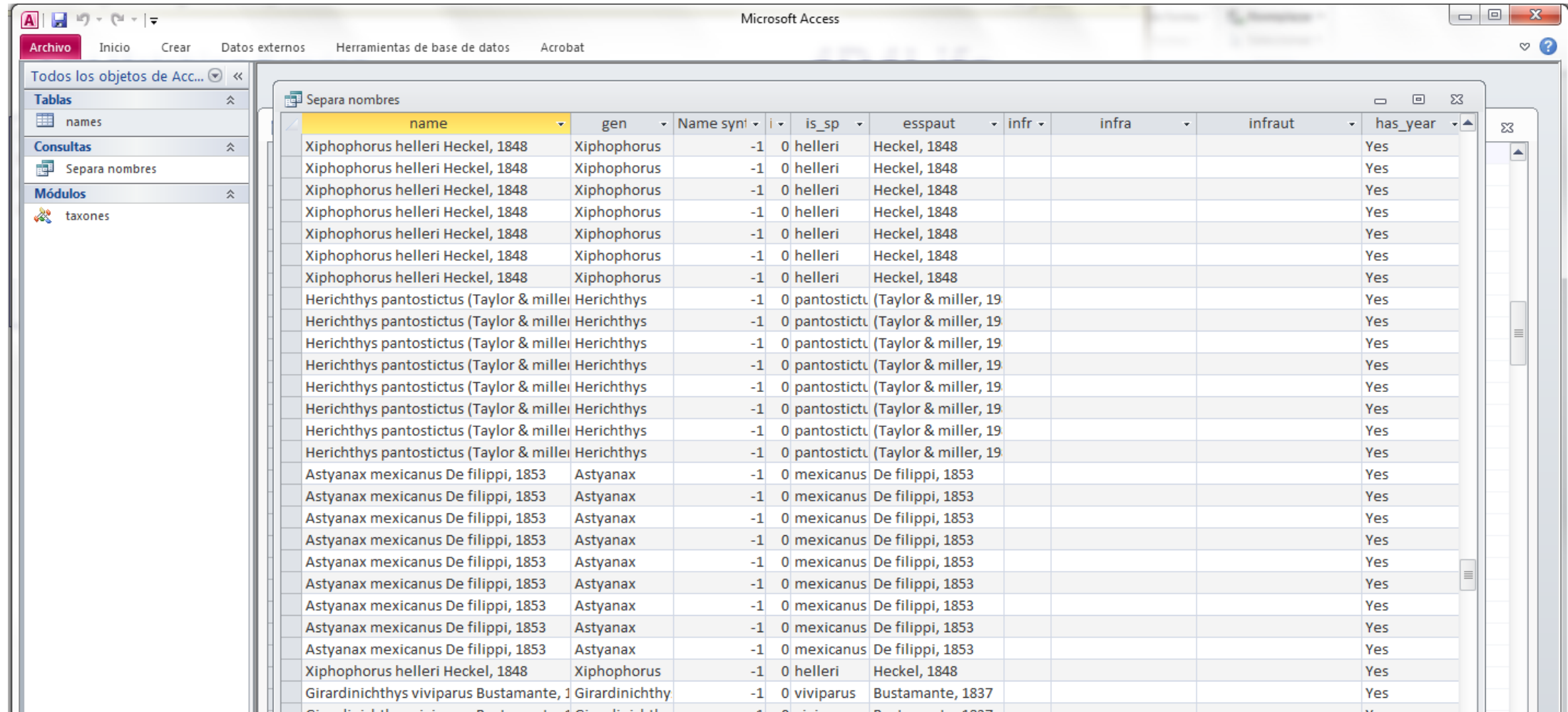
Worms Taxon Match (especies marinas)



Dominio taxonómico

Atomización de nombres

NAME
parser



Microsoft Access

Archivo Inicio Crear Datos externos Herramientas de base de datos Acrobat

Todos los objetos de Acc... <<

Tablas

- names

Consultas

- Separa nombres

Módulos

- taxones














Separa nombres

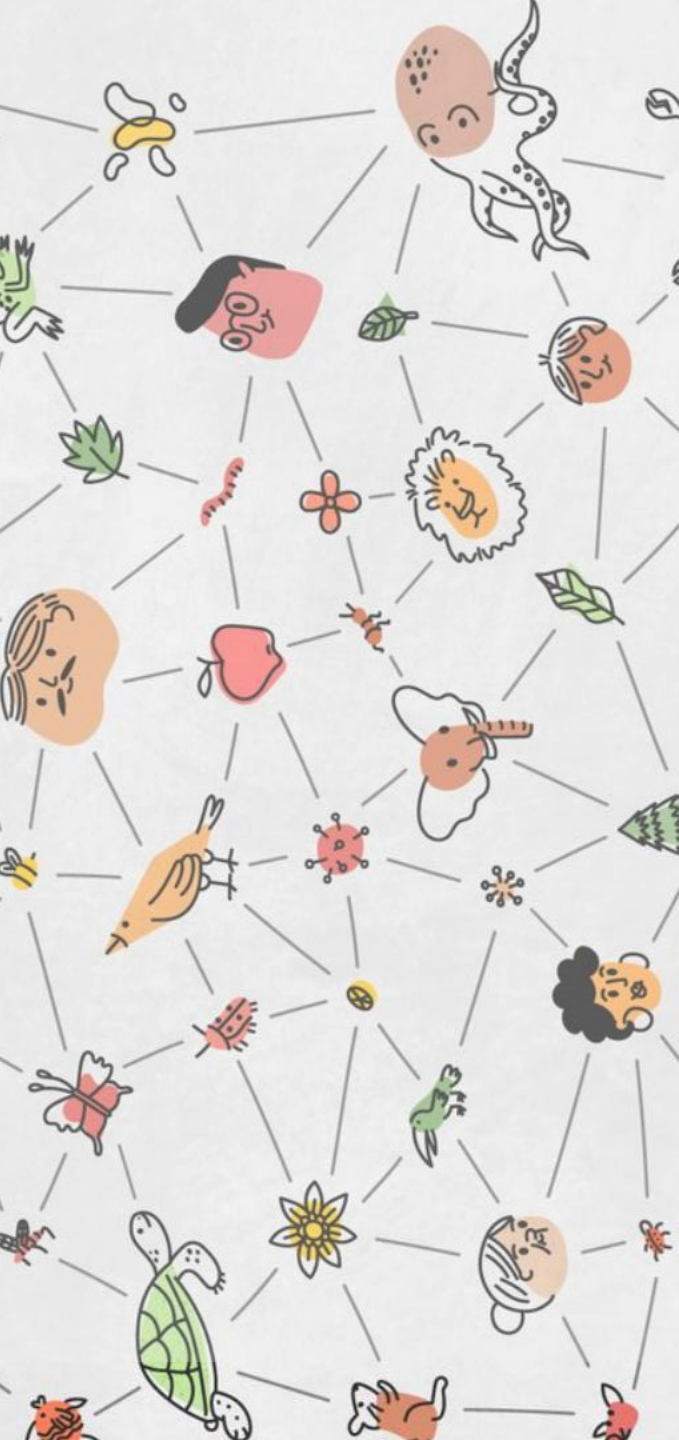
name	gen	Name syn	i	is_sp	esspaut	infr	infra	infraut	has_year
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Herichthys pantostictus (Taylor & miller	Herichthys	-1	0	pantostictu	(Taylor & miller, 19				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Astyanax mexicanus De filippi, 1853	Astyanax	-1	0	mexicanus	De filippi, 1853				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Girardinichthys viviparus Bustamante, 1	Girardinichthy	-1	0	viviparus	Bustamante, 1837				Yes

Dominio taxonómico

Resolución



<div>Get dataShareToolsInside GBIF<div>katia</div></div>								
TOOLS LOOK UP								
OriginalName	PreferedKingdom	MatchType	Confidence	ScientificName (Editable)	Status	Rank	Kingdom	Phylum
Atrichum undulatum (Hedw.) P. Beauv.	any	EXACT	100	 Atrichum undulatum Palisot de Beauvois, 1805	ACCEPTED	species	Plantae	Bryophyta
Aulacomnium androgynum (Hedw.) Schwaegr.	any	EXACT	100	 Aulacomnium androgynum Schwaegrichen, 1827	ACCEPTED	species	Plantae	Bryophyta
Aulacomnium palustre (Hedw.) Schwaegr.	any	EXACT	100	 Aulacomnium palustre Schwaegrichen, 1827	ACCEPTED	species	Plantae	Bryophyta
Barbilophozia kunzeana (Huebener) Müll. Frib.	any	EXACT	97	 Barbilophozia kunzeana (Huebener) Müll.Frib.	ACCEPTED	species	Plantae	Marchantioph
Barbula bolleana	any	EXACT	98	 Barbula bolleana Brotherus, 1924	ACCEPTED	species	Plantae	Bryophyta
Barbula convoluta Hedw.	any	EXACT	100	 Barbula convoluta Hedwig, 1801	ACCEPTED	species	Plantae	Bryophyta
Barbula convoluta Hedw. var. sardoa Bruch & Schimp	any	EXACT	100	 Barbula convoluta var. sardoa Schimp.	ACCEPTED	variety	Plantae	Bryophyta
Barbula unguiculata Hedw.	any	EXACT	100	 Barbula unguiculata Hedwig, 1801	ACCEPTED	species	Plantae	Bryophyta
Brachythecium albicans (Hedw.) Schimp.	any	EXACT	100	 Brachythecium albicans W. P. Schimper in B.S.G., 1853	ACCEPTED	species	Plantae	Bryophyta
Brachythecium dieckii Roll	any	FUZZY	99	 Brachythecium dieckeii Röll, 1897	ACCEPTED	species	Plantae	Bryophyta



Herramientas de validación

Herramientas de validación

Herramientas específicas que comprueban la calidad o validez de los datos para su publicación una vez que estos ya están adaptados al estándar.



Darwin Test




DATA VALIDATOR

Herramientas de validación



DARWIN_TEST: Validación de datos exportados a GBIF con Darwincorev2

 **Darwin Test**
Validador de calidad de datos



Data validation and geographic coordinates generalization for Darwin Core datasets <http://www.gbif.es/>

1. Seleccionar

Tipo de datos a validar:

☒ DarwinCore 1.2 (mdb) ☐ DarwinCore 1.4 (mdb) ☐ DarwinCore Archive (zip)

Seleccione el origen de los datos a validar

C:\Darwintest_3.4\ColeccionesConCaracteresAnomalos\DarwinCore2_UPNA-H_20171106_GBIF.mdb ... Aceptar

2. Validación de datos

2.1. Metadatos

Detectar posibles errores

↓

Visualización y chequeo de datos de DwC

↗

↘

Actualizar tabla DwC vinculada

Crear tabla DwC para corregir base de datos original ?

Herramientas de validación



Data validator

Get dataShareToolsInside GBIF

NEW VALIDATION

SUMMARYDARWIN CORE EXTENSIONS

● The file can be indexed by GBIF

Some issues were detected by the validator:

GBIF Occurrence Interpretation

Recorded date mismatchTaxon match fuzzyTaxon match higherrankCoordinate reprojected

This report has been written to <https://www.gbif.org/tools/data-validator/1509439657203> It was generated in 10 minutes and will be deleted after one month. Until then you can revisit the report at your convenience.

File Format: Tabular File (.csv, .tsv)
Media Type: text/plain
Core Row Type: Darwin Core Occurrence
Extensions: 0

Core

CARIMED_2016_OCCURRENCE.TXT

- Term Frequency
- Validation Issues

CARIMED_2016_Occurrence.txt

Row Type: [Darwin Core Occurrence](#)

Number Of Lines: 5.979

Number Of Rows With Interpreted Taxon: 5.979

Number Of Interpreted Dates: 5.979

Number Of Interpreted Coordinates: 5.979

Term Frequency

Term	Count	Percentage	Interpreted
dwc:eventID	5.979	100%	
dcterms:type	5.979	100%	
dcterms:modified	5.979	100%	

<https://www.gbif.org/tools/data-validator>

Lo que nos gustaría

Darwin Core Translator

⌘A Texto

📄 Documentos

Mi base de datos

ESPAÑOL

FRANCÉS

↔

Darwin Core

CÉS

🎤

0/5000

Enviar comentarios

Lo que se le parece



gbif-norway/
ChatIPT



Aa 3

Contributors

3

Issues

9

Stars

1

Fork



<https://chatipt.svc.gbif.no/>

<https://github.com/gbif-norway/ChatIPT/activity>

Cerca de DarwinCore translator

ChatIPT



<https://chatipt.svc.gbif.no/>

Es un asistente conversacional (como ChatGPT) especializado en ayudarte a transformar las hojas de cálculo en archivos listos para publicar bajo el formato Darwin Core, utilizado en portales como GBIF.



Sign in to ORCID

Don't have your ORCID iD yet? [Register now](#)

Email *or* ORCID iD

0000-0003-3077-6136

For example: joe@institution.edu or 0000-1234-5678-9101

Password

.....

Cerca de DarwinCore translator

ChatIPT



<https://chatipt.svc.gbif.no/>

Welcome to ChatIPT

ChatIPT is a chatbot for students and researchers who are new to data publication or only occasionally publish data.

It cleans and standardises spreadsheets, creates basic metadata, asking the user for guidance where necessary through natural conversation. Finally, it publishes the data on gbif.org as a Darwin Core Archive.

Why is it necessary?

At a conservative estimate, there are 300 - 400 PhDs and MScs in Europe alone at the moment generating biodiversity data, as well as countless small biodiversity research studies.

Cerca de DarwinCore translator

ChatIPT



<https://chatipt.svc.gbif.no/>

1. Verifica si el archivo contiene datos publicables
2. Separa sub-tablas si las hay
3. Elimina filas o columnas innecesarias (como totales, comentarios, etc.)
4. ChatIPT revisa tus datos y te propone cuál "Core" usar (cecklist, occurrences, sampling event) y puede sugerirte extensiones.
5. Convierte los datos al formato Darwin Core (fechas, coordenadas, alturas...)
6. Descarga los archivos

Cerca de DarwinCore translator

ChatIPT



<https://chatipt.svc.gbif.no/>

Butterfly observations in Sierra Nevada, Spain (2016–2020)

Started 10/29/2025, 8:00:55 AM

Structure Notes

Visual counts of butterflies recorded in the Sierra Nevada National Park, Granada, Spain between 2016 and 2020. Each record includes sampling event information, quantity of individuals, elevation, habitat description and taxonomic identification.

Task: Data structure exploration - (Table ID 206) complete

Show code + results

Show code results

Show code results

Show code results

Show code results

Datos originales (ID 206)

id institution	código de la institución	código de
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		
4b4b2111-ee51-45f5-bf5e-f5...		

Working... loading next task

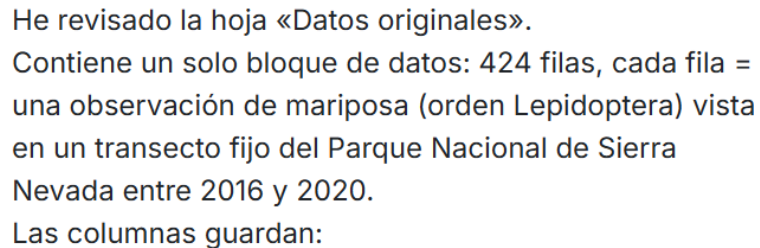
ChatIPT



Started 10/29/2025, 8:00:55 AM

Task: Data structure exploration - (Table ID 206) **complete** ☒

Task: Data content exploration - (Table ID 206)



- id evento → visita de muestreo
- fecha, localidad, elevación, hábitat, coordenadas, datum
- id registro → observación única

Datos originales (ID 206)

[illegible]

Cerca de DarwinCore translator

ChatIPT

<https://chatipt.svc.gbif.no/>



qué tengo que hacer?



Right now you only need to tell me one thing:

`institutionCode`

This is a short tag (3-10 letters) for the institute that owns or will host the data.

Examples: "MNCN", "EBD-CSIC", "OBSNEV".

If you already use a code, please give it to me.

If you are not sure, just write the full institute name and I will suggest a code.

institution code=OBSNEV



Show code results



Great, I'll use `institutionCode = "OBSNEV"`.

Next, we also need `collectionCode` – a short tag for this particular butterfly monitoring dataset (for example "BUTTERFLIES", "TransectsSN", or similar).

If you already have one, please let me know.

If not, just suggest a short name (no spaces) and we'll use that.

BUTTERFLIES

Working...



ChatIPT

 ChatIPT[illegible]



¡Gracias!

katia@gbif.es

www.gbif.es



MINISTERIO
DE CIENCIA
E INNOVACIÓN



CSIC



Sociedad Ibérica para la Conservación y
Estudio de los Mamíferos



GRUPO DE
ESPECIALISTAS DE
ESPECIES ESPAÑA

