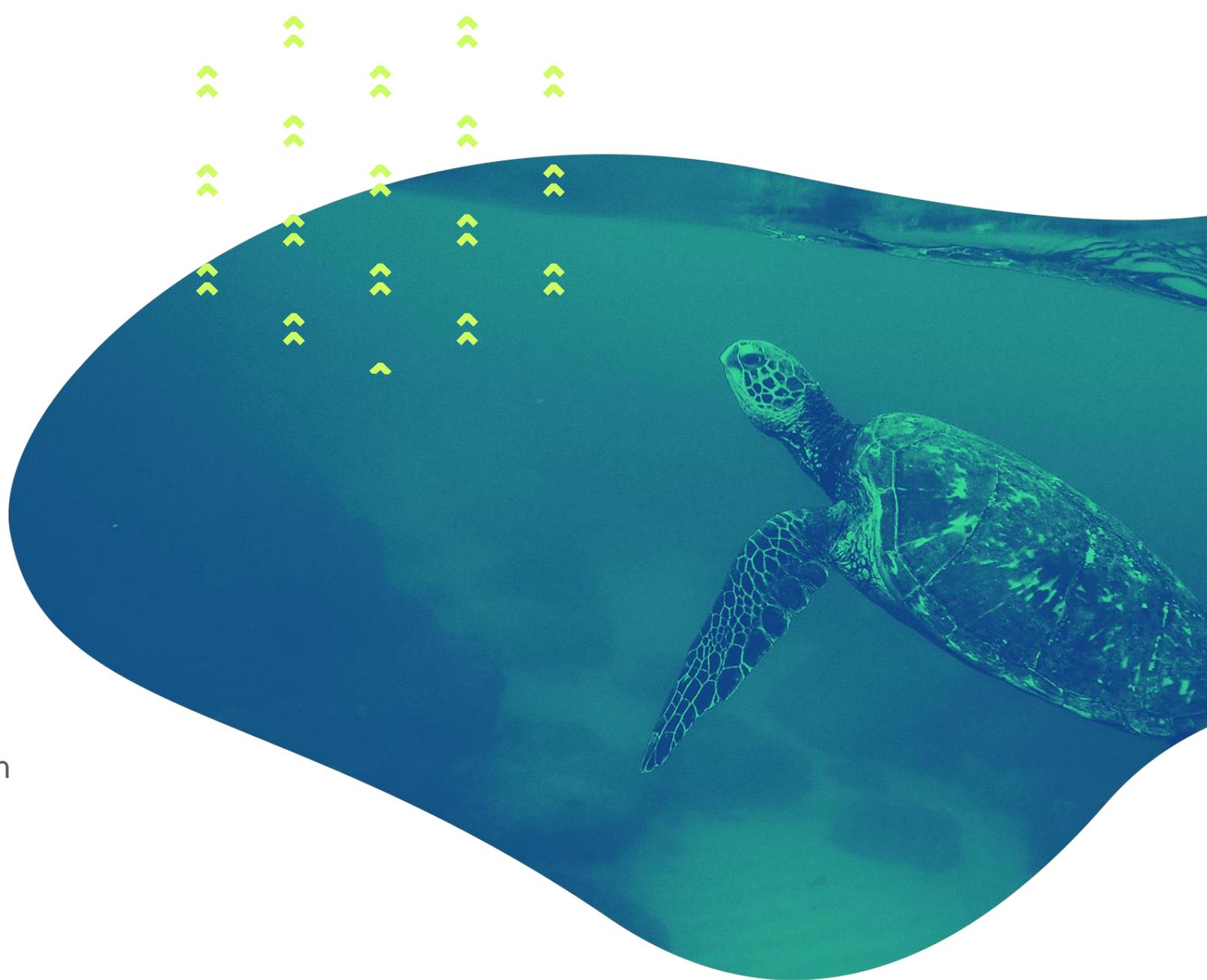


Herramientas útiles para la calidad y limpieza de datos

Publicación de datos de biodiversidad en
GBIF y en revistas científicas

Katia Cezón
katia.gbif@gmail.com

Gbif, Es



II Taller GBIF.ES: Publicación de datos de biodiversidad en GBIF y en revistas científicas



Herramientas útiles para la calidad y limpieza de datos

Dominios

- **Geo-espacial:** Datos con coordenadas o georreferenciados y regiones político-administrativas documentadas.
- **Taxonomía (nomenclatura):** Nombres científicos, niveles taxonómicos
- **Formato (estandarización):** Errores de tipeo, formatos de fecha, formatos de coordenadas, caracteres especiales y demás.

Aspectos a tener en cuenta

- Precio, disponibilidad, licencia
- Requerimientos técnicos
- Facilidad de uso
- Documentación y soporte
- Flexibilidad
- Automatización
- Etc.



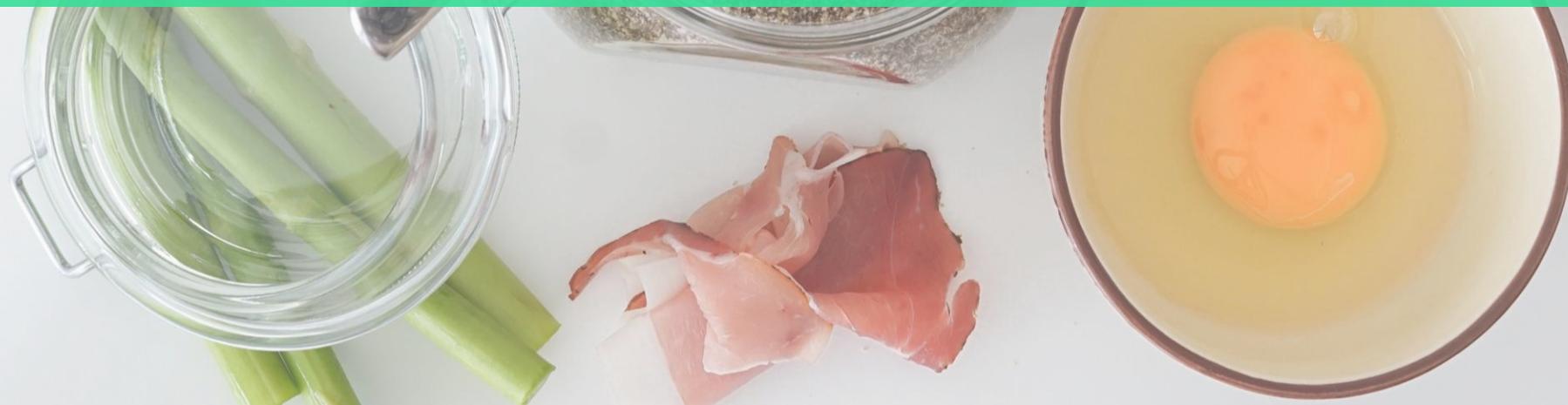
Aspectos a tener en cuenta

- 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





No existe una receta

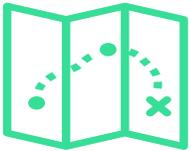


Herramientas útiles para la calidad y limpieza de datos



Herramientas genéricas para el almacenamiento y gestión de datos

- Excel, Access, Open Office, etc.



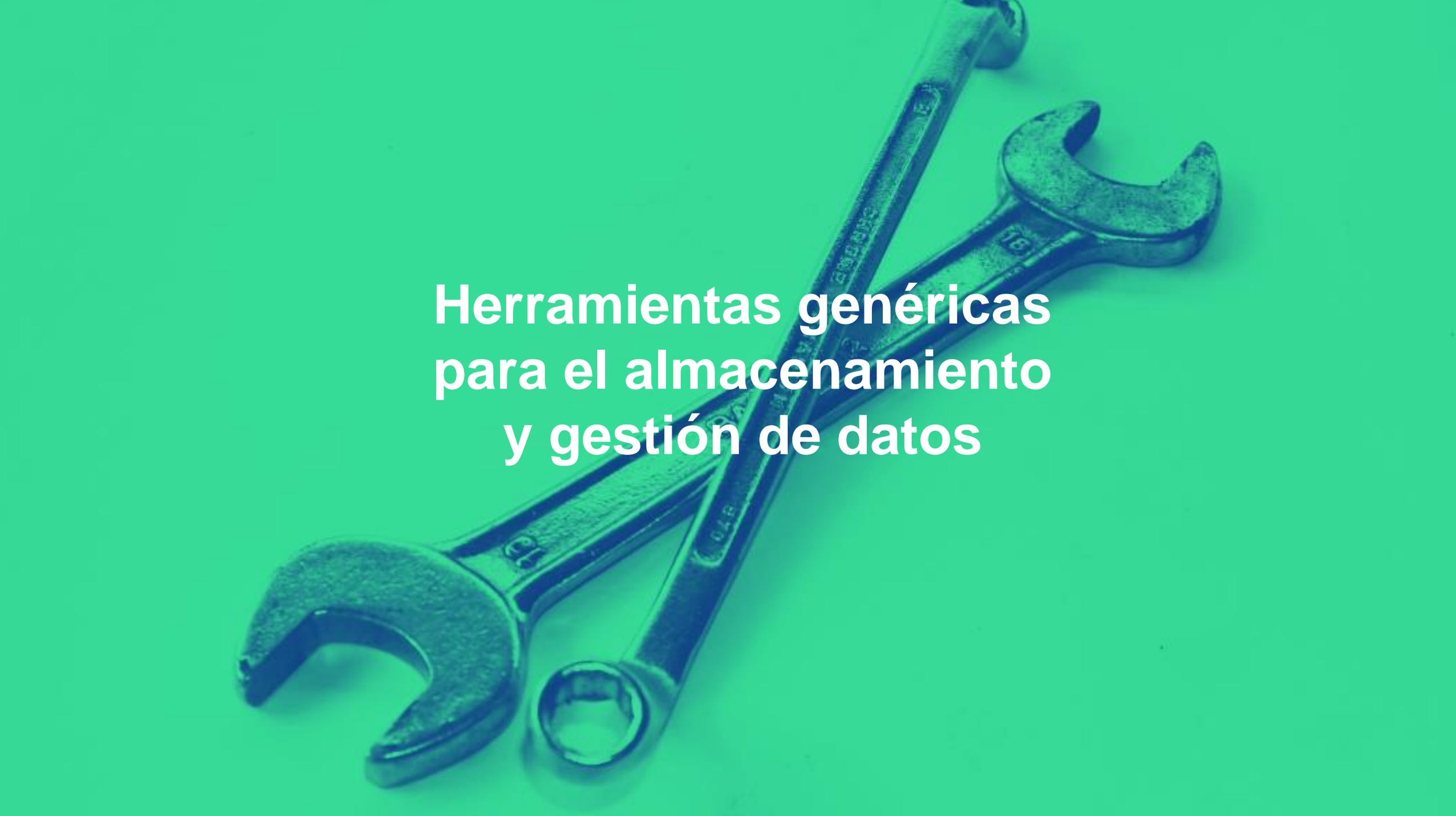
Herramientas para el tratamiento de nombres científicos, fechas y coordenadas

- Herramientas para la gestión de nombres científicos (atomización, herramientas para comprobar status, búsqueda de autores, etc.)
- Herramientas geográficas (visualización, comprobación de coordenadas, conversión, etc.)
- Herramientas para el tratamiento de las fechas



Herramientas para la limpieza y validación de datos

- Open Refine
- Darwin Test
- Darwin Core Archive Validator

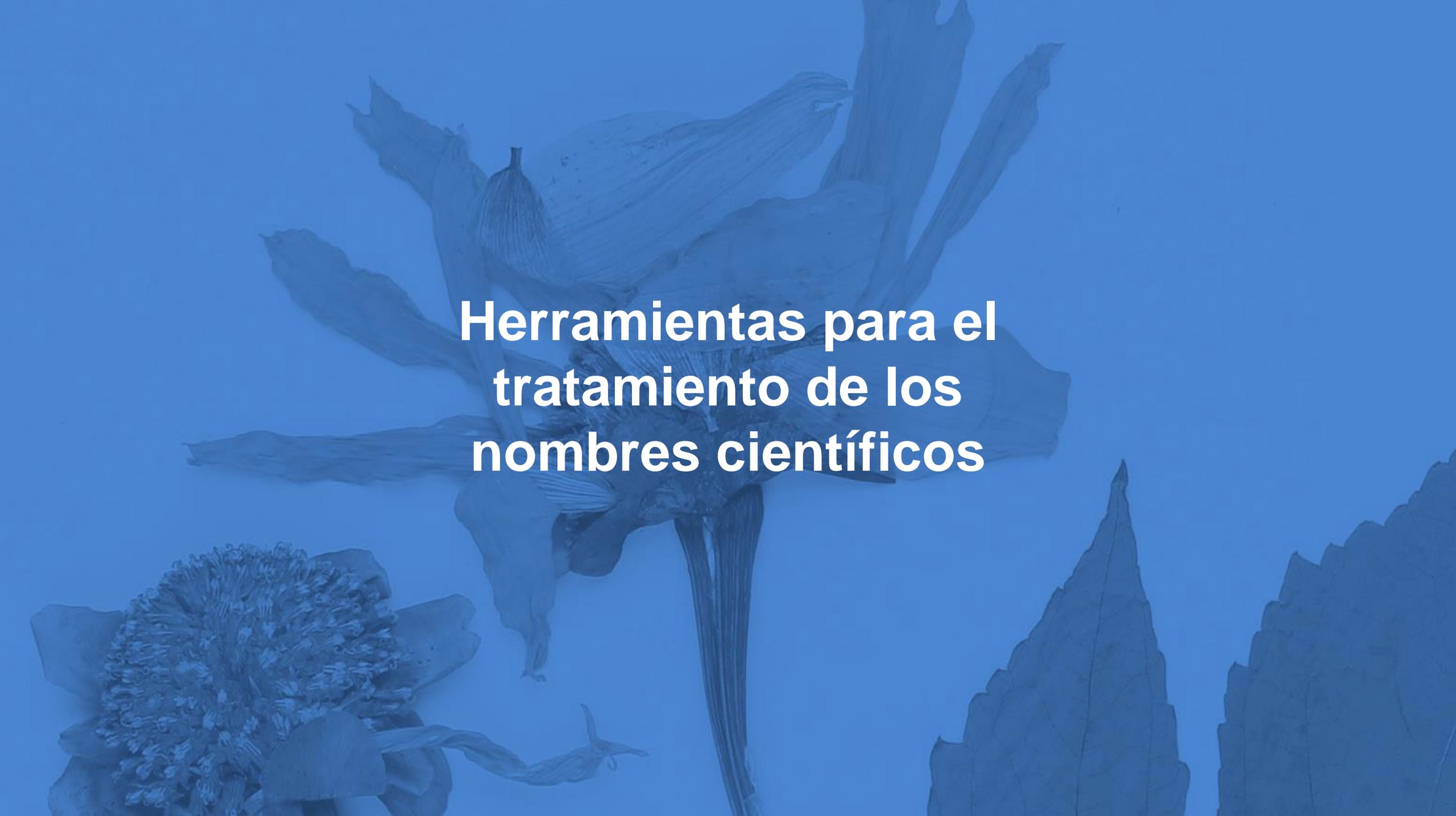
The image features three crossed wrenches on a plain white background. The wrenches are arranged in a 'V' shape, with one at the top and two at the bottom. They are made of metal and show signs of use. The text is centered over the wrenches.

**Herramientas genéricas
para el almacenamiento
y gestión de datos**

Herramientas de almacenamiento y gestión de datos

Permiten manejar los datos mediante tablas (formadas por filas o registros y columnas o variables), crear relaciones entre tablas, elaborar consultas y formularios para introducir datos o informes para extraer la información.



A botanical illustration of a flower and its parts, including a large flower with a prominent central cluster of stamens, a smaller flower, and several leaves, all rendered in a light blue color against a darker blue background. The text is overlaid on the central flower.

Herramientas para el tratamiento de los nombres científicos

Herramientas para el tratamiento de los nombres científicos

Permiten realizar tareas como la **atomización** de sus componentes (género, epíteto específico, etc.), permiten la **resolución de nombres científicos** de cualquier grupo taxonómico, generan la jerarquía taxonómica, etc.

Atomización

[GBIF - Name parser](#)

[Name Parser GBIF España](#)

Resolución

[List Matching Service](#)

[Global Names Resolver](#)

[T-REX](#)

[iPlant](#)

[Species matching GBIF](#)

Herramientas para el tratamiento de los nombres científicos



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 syni	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
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
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

Herramientas para el tratamiento de los nombres científicos



Matched Species

[Download to file](#)

Your Data	Scientific Name	Status
Helianthemum squamatum (L.) Pers.	<i>Helianthemum squamatum</i> (L.) Pers.	accepted name
Thymus lacaitae Pau	<i>Thymus lacaitae</i> Pau	accepted name
Thymus lacaitae Pau	<i>Thymus lacaitae</i> Pau	accepted name
Thymus lacaitae Pau	<i>Thymus lacaitae</i> Pau	accepted name
Thymus lacaitae Pau	<i>Thymus lacaitae</i> Pau	accepted name
Thymus vulgaris L.	<i>Thymus vulgaris</i> L.	accepted name
Thymus vulgaris L.	<i>Thymus vulgaris</i> L.	accepted name
Lepidium subulatum L.	<i>Lepidium subulatum</i> L.	accepted name
Lepidium subulatum L.	<i>Lepidium subulatum</i> L.	accepted name
Lepidium subulatum L.	<i>Lepidium subulatum</i> L.	accepted name
Lepidium subulatum L.	<i>Lepidium subulatum</i> L.	accepted name
Lepidium subulatum L.	<i>Lepidium subulatum</i> L.	accepted name
Centaurea hyssopifolia Vahl	<i>Centaurea hyssopifolia</i> Vahl	accepted name
Centaurea hyssopifolia Vahl	<i>Centaurea hyssopifolia</i> Georgi	synonym

Herramientas para el tratamiento de los nombres científicos



Resultados de la búsqueda taxonómica

Descargar

XLSX

Fuentes seleccionadas: GBIF Backbone Taxonomy

Total Datos 19

Página Actual 1 de 2

Anterior

Siguiente

Nombre ingresado	Nombre científico resuelto	Fuente de información	Puntaje de coincidencia	Tipo de Coincidencia	Detalles
Herichthys pantostictus (Taylor & miller, 1983)	Nosferatu pantostictus (Taylor & Miller, 1983)	GBIF Backbone Taxonomy	0.999	Coincidencia exacta	Detalles
Agonostomus monticola Bancroft, 1834	Agonostomus monticola (Bancroft, 1834)	GBIF Backbone Taxonomy	0.997	Coincidencia exacta del nombre canónico	Detalles

Herramientas para el tratamiento de los nombres científicos



Species
matching



TOOLS | LOOK UP

Normalize species names from a csv file against the GBIF backbone.
The file is expected to be have a column called 'scientificName' and an optional column 'kingdom' and 'id'.

SIMPLEEXAMPLE.CSV

ADVANCEDEXAMPLE.CSV

SELECT FILE

or

DROP HERE

<https://www.gbif.org/tools/species-lookup>

Herramientas para el tratamiento de los nombres científicos



Species
matching



Get data

Share

Tools

Inside GBIF



katia

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 dieckei Röll, 1897	ACCEPTED	species	Plantae	Bryophyta

<https://www.gbif.org/tools/species-lookup>

A person in a white coat stands on a map of Europe, pointing upwards. The map is overlaid with a green and blue gradient. The text "Herramientas para el tratamiento de los datos geográficos" is centered over the map.

Herramientas para el tratamiento de los datos geográficos

Herramientas para el tratamiento de los datos geográficos

Permiten visualizar puntos en mapas, comprobar coordenadas o convertirlas al formato necesario para la publicación, etc.

Comprobación

[Info XY \(species link tools\)](#)

[Excel to kml \(Earth point\)](#)

[Google Earth, Google Maps,](#)

[Carto](#)

Sistemas de información geográfica

Conversión

[Canadensys coordinates conversion](#)

Conversor de coordenadas GBIF.ES

[Geotrans](#)

Herramientas para el tratamiento de los datos geográficos

id , longitude , latitude (decimal degree)

```
Jilotla -98.741583 20.551972
Jilotla -98.741583 20.551972
Jilotla -98.741583 20.551972
Jilotla -98.741583 20.551972
Jilotla -98.741583 20.551972
Jilotla -98.741583 20.551972
Jihuico -98.727305 20.541722
Jihuico -98.727305 20.541722
Jihuico -98.727305 20.541722
Jihuico -98.727305 20.541722
```

output:

HTML

see map

Search

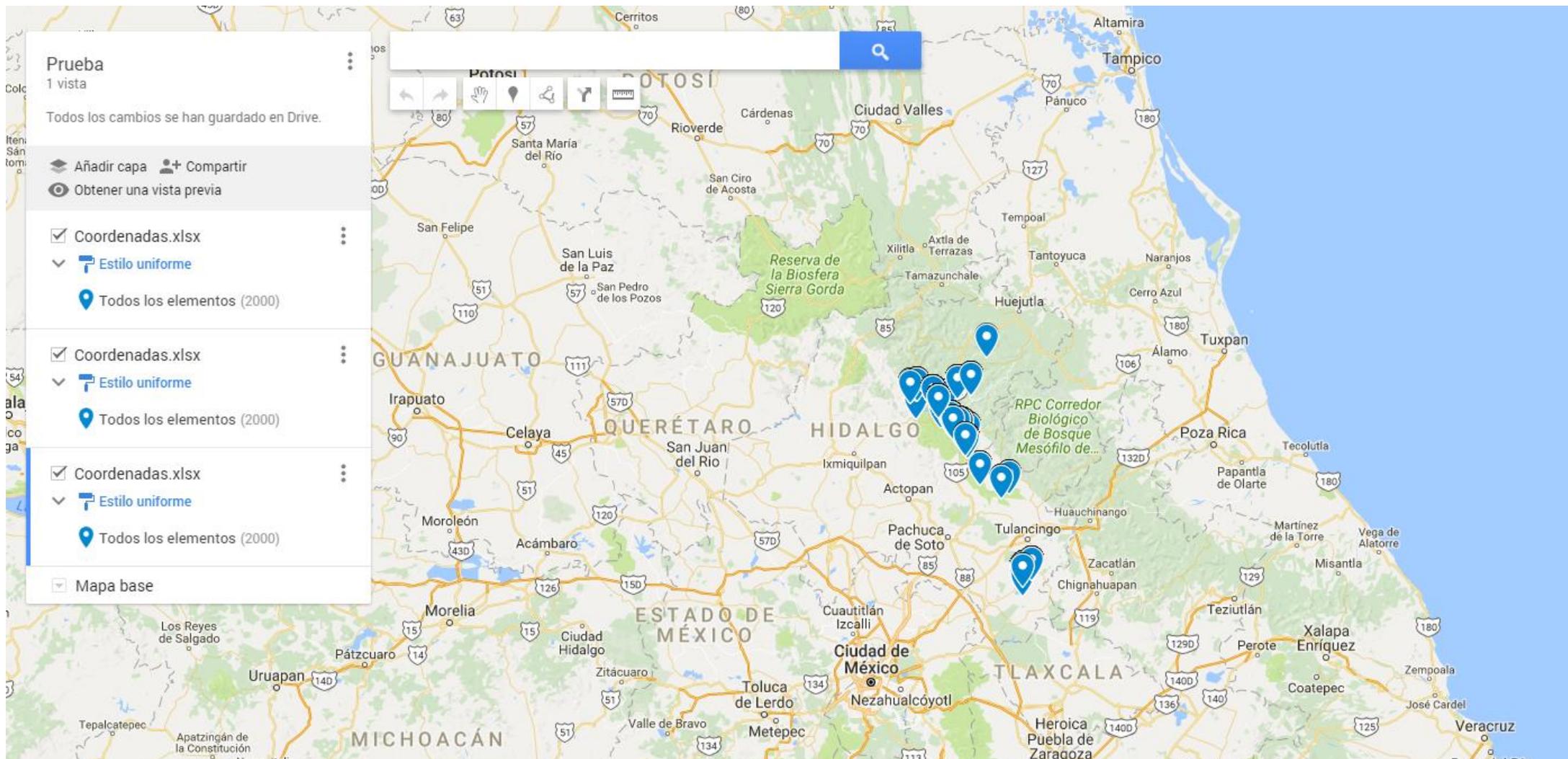
Results

id	longitude	latitude	country	admin1	typeadmin1	admin2	typeadmin2	admin3	typeadmin3	admin4	typeadmin4
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jilotla	-98.741583	20.551972	México	Hidalgo	State	Metztitlán	Municipality				
Jihuico	-98.727305	20.541722	México	Hidalgo	State	Metztitlán	Municipality				
Jihuico	-98.727305	20.541722	México	Hidalgo	State	Metztitlán	Municipality				



Google Maps

Herramientas para el tratamiento de los datos geográficos





Coordinate conversion

Use this tool to convert geographic coordinates from DDMSS to decimal degrees. Type coordinate pairs on separate lines or paste latitude and longitude columns from a spreadsheet. Each row may be optionally preceded by an identifier followed by a pipe or tab.

45° 32' 25" N, 129° 40' 31" W

Submit

[Coordinate conversion](#)

[Date parsing](#)

[Tools API](#)

[About](#)

Example input

45° 32' 25" N, 129° 40' 31" W

1 | 45.5° N, 129.6° W

2 | 40°26'47"N,74° 0' 21.5022"W

feedback



Herramientas para el tratamiento de los datos geográficos

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

Archivo Inicio Crear Datos externos Herramientas de base de datos Acrobat Herramientas de tabla Campos Tabla

Todos los objetos de Acc... << >>

Buscar...

Tablas

- Coordenadas

Consultas

- 0_Convierte coordenadas

Módulos

CatalogNumber	Coordenadas	coordinateUr	decimalLati	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

Herramientas para el tratamiento de las fechas



Herramientas para el tratamiento de las fechas

Herramientas que permiten la unificación y transformación de distintos formatos de fechas.

Atomización y unificación de fechas

[Canadensys date parsing](#)

Función “fechasa” disponible en diversas aplicaciones de GBIF.es

13-VI-1980 → 1980-06-13

13 Junio 1980 → 1980-06-13

13-06-1980 → 1980-06-13



Herramientas para el tratamiento de las fechas

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

- [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



**Herramientas de
depuración y
validación para la
publicación**

Herramientas de validación y depuración para la publicación

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 y depuración para la publicación

DARWIN_TEST: Validación de datos exportados a GBIF con Darwincorev2



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 ↗

Detectar caracteres ASCII anómalos ↘

Actualizar tabla DwC vinculada

Crear tabla DwC para corregir base de datos original ?

Herramientas de validación y depuración para la publicación



Data validator



Get data

Share

Tools

Inside GBIF



katia

TOOLS | DATA VALIDATOR

This an early access version. Please report issues and feedback [here](#).

NEW DATA VALIDATION

ABOUT

SELECT FILE

or

DROP HERE

or Fetch file from location:

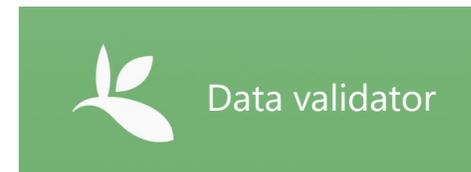
<http://example.com/dwca.zip>

SUBMIT

File size limit: 100 mb

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

Herramientas de validación y depuración para la publicación



SUMMARY

DARWIN CORE EXTENSIONS

NEW VALIDATION

The file can be indexed by GBIF

Some issues were detected by the validator:

GBIF Occurrence Interpretation

- Recorded date mismatch
- Taxon match fuzzy
- Taxon match higherrank
- Coordinate reprojected

File Format: Tabular File (.csv, .tsv)

Media Type: text/plain

Core Row Type: Darwin Core Occurrence

Extensions: 0

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.

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%	

Gbif.es

¡Gracias!

pzermoglio@gmail.com

Gbif.es

VerNe



Universidad
de Navarra