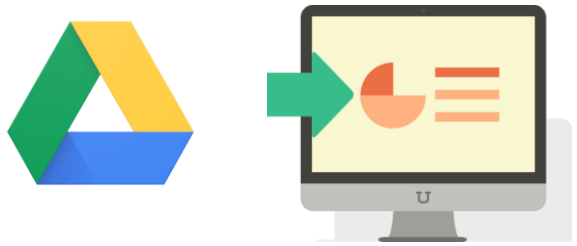
Strengthening Zimbabwe's GBIF node through collaboration with GBIF Spain

CESP PROJECT. 24-27 SEPTEMBER 2019. MADRID SPAIN

Work materials

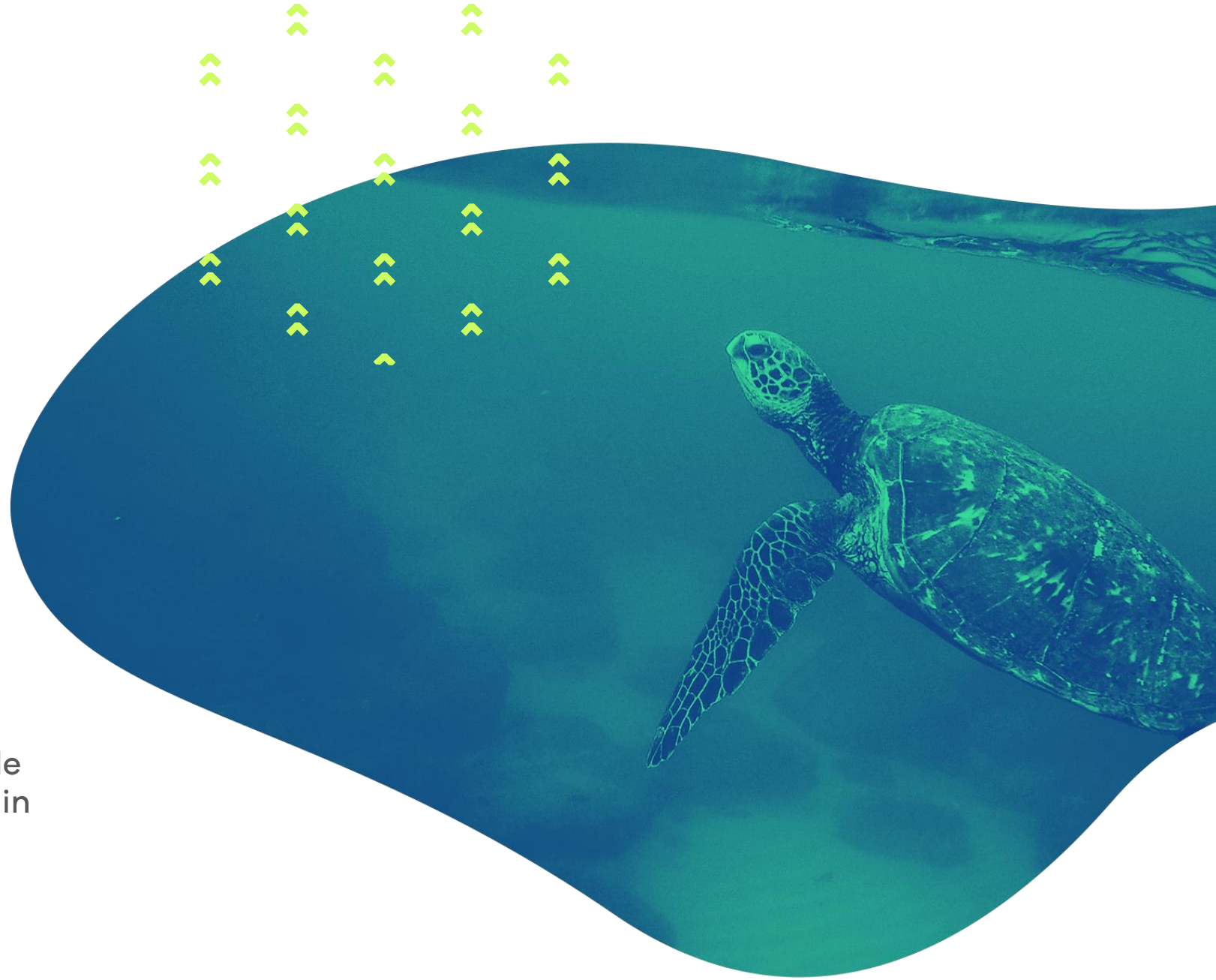
The workshop agenda and materials are online accessible at the following…



…*Google Drive* link

https://bit.ly/2mor5lp

Useful tools for data quality and cleansing



Strengthening Zimbabwe's GBIF node through collaboration with GBIF Spain

Miguel Vega
miguel.vega@rjb.csic.es

Gbif.es

Contents

1. Previous considerations
2. Types of tools:
    - Data storage and management
    - Related to scientific names
    - Related to dates
    - Related to geographical data
    - Data validation before publishing
3. Demos
4. Use cases

Previous considerations

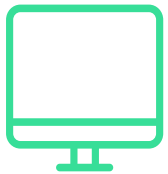
- Prize, accessibility, licence
- Technical requirements
- Simplicity when using
- Documentation and support
- Flexibility
- Automation
- Etc.

Previous considerations

- It’s very important that the person responsible for data management has a **good understanding of data handling** because s/he will need to make corrections, migrate from/to different formats, etc.

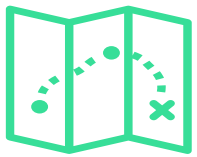
There is not a perfect recipe

Useful tools for data quality



General tools for data storage and management

- Excel, Access, Open Office, etc.



Tools for the validation (and cconvertion) of scientific names, dates and coordinates

- Tools for scientific name management (status, authors, taxa clasification, etc.)
- Geographical tools (coordinates visualization and conversion)
- Tools for data conversion

Tools for data validation

- Open Refine
- Darwin Test
- Darwin Core Archive Validator

General tools for data storage and management

General tools for data storage and management

They allow the data management through the use of spread sheets and other functionalities hosted in common software like OpenOffice, Ms. Excel, Ms. Access, SQLite, etc.

Gbif.es

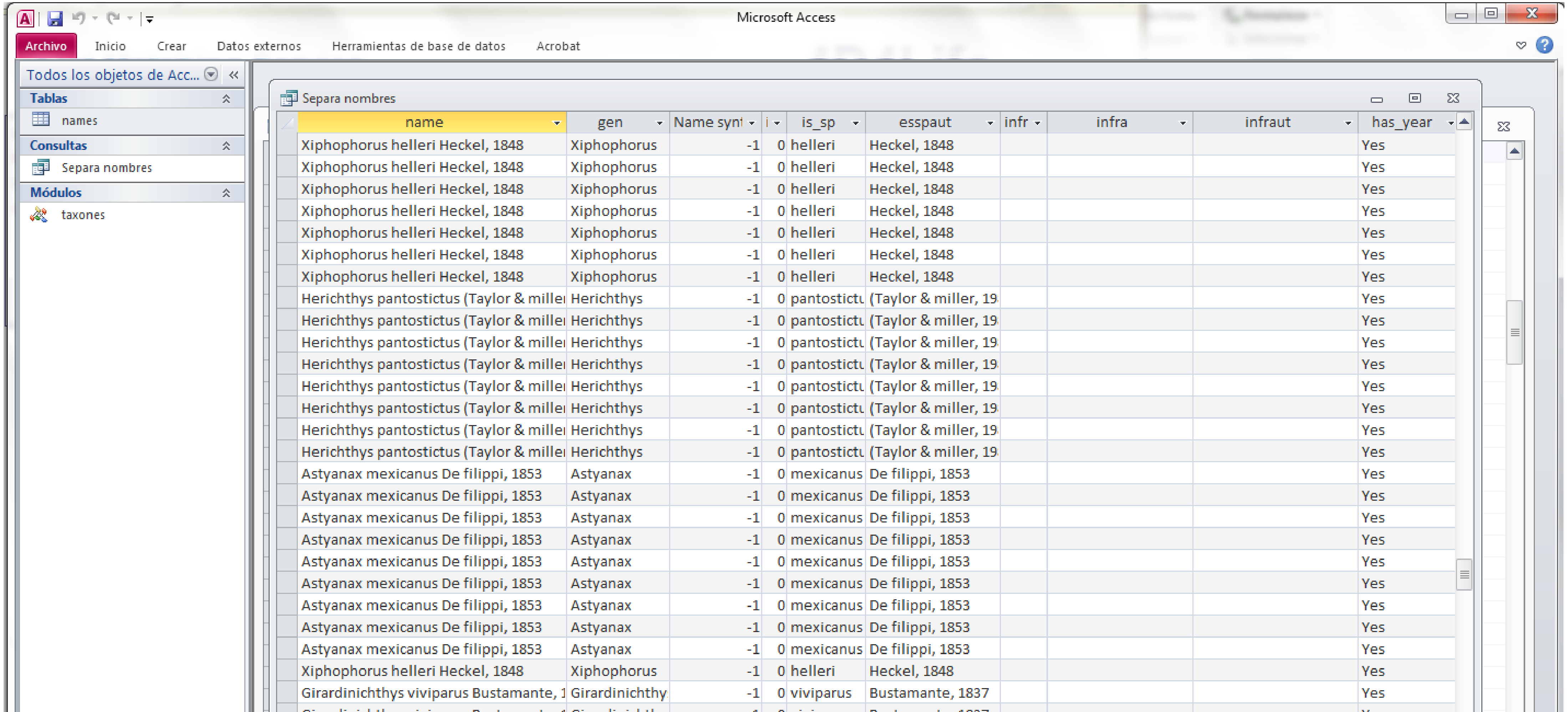
Tools for managing scientific names

Tools for managing scientific names

They allow to **parse scientific names** in their basic components (genus, specific epithet,...), to **validate used scientific names** from every taxonomic group, to establish a taxonomic rank, etc.

Parsing	GBIF - Name parser	**Validation**	List Matching Service
	Name Parser GBIF Spain		Global Names Resolver
			T-REX
			iPlant
			Species matching GBIF

Tools for managing scientific names



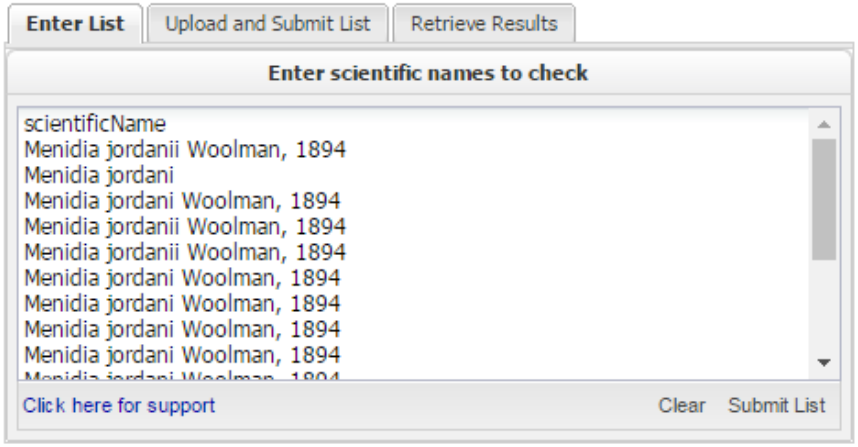
name	gen	Name synt	i	is_sp	esspaut	infr	infra	infraut	has_year
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Xiphophorus 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
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
Xiphophorus helleri Heckel, 1848	Xiphophorus	-1	0	helleri	Heckel, 1848				Yes
Girardinichthys viviparus Bustamante, 1	Girardinichthy	-1	0	viviparus	Bustamante, 1837				Yes

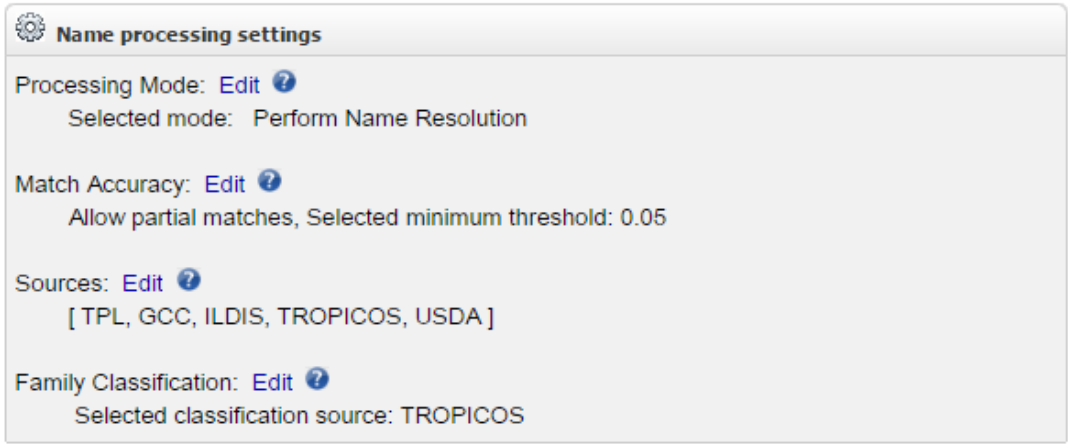
Tools for managing scientific names

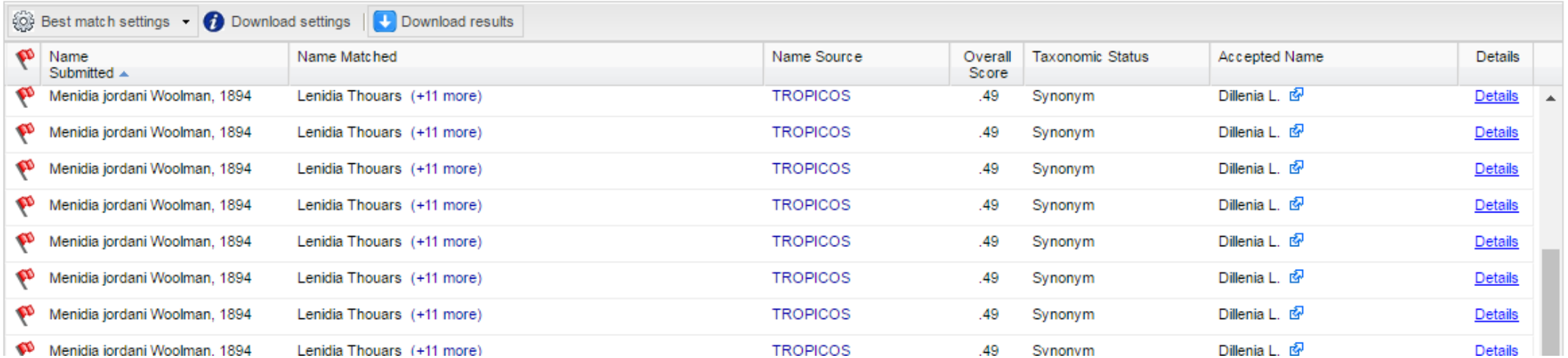
Matched Species Download to file

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

Tools for managing scientific names

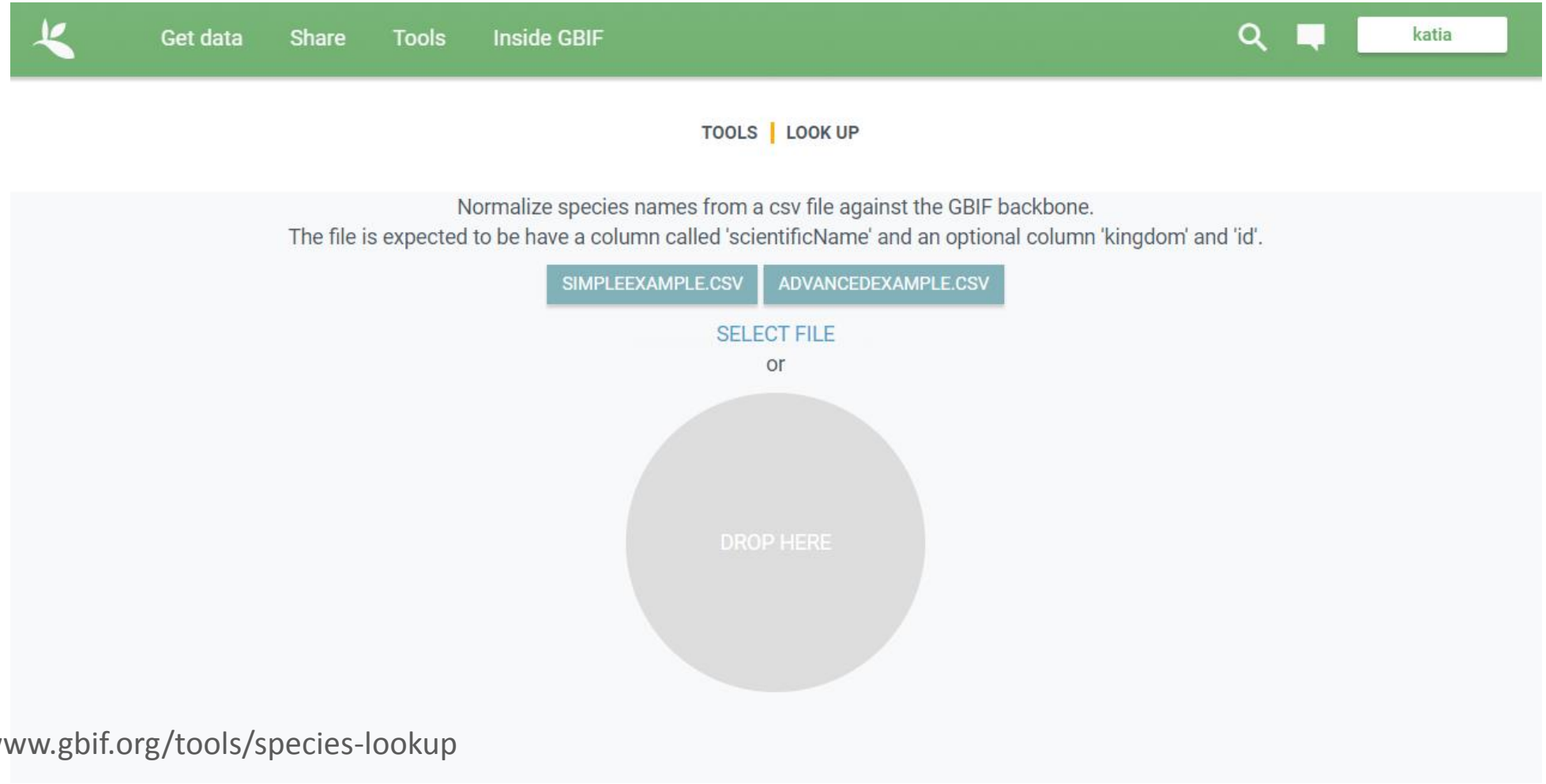






Tools for managing scientific names

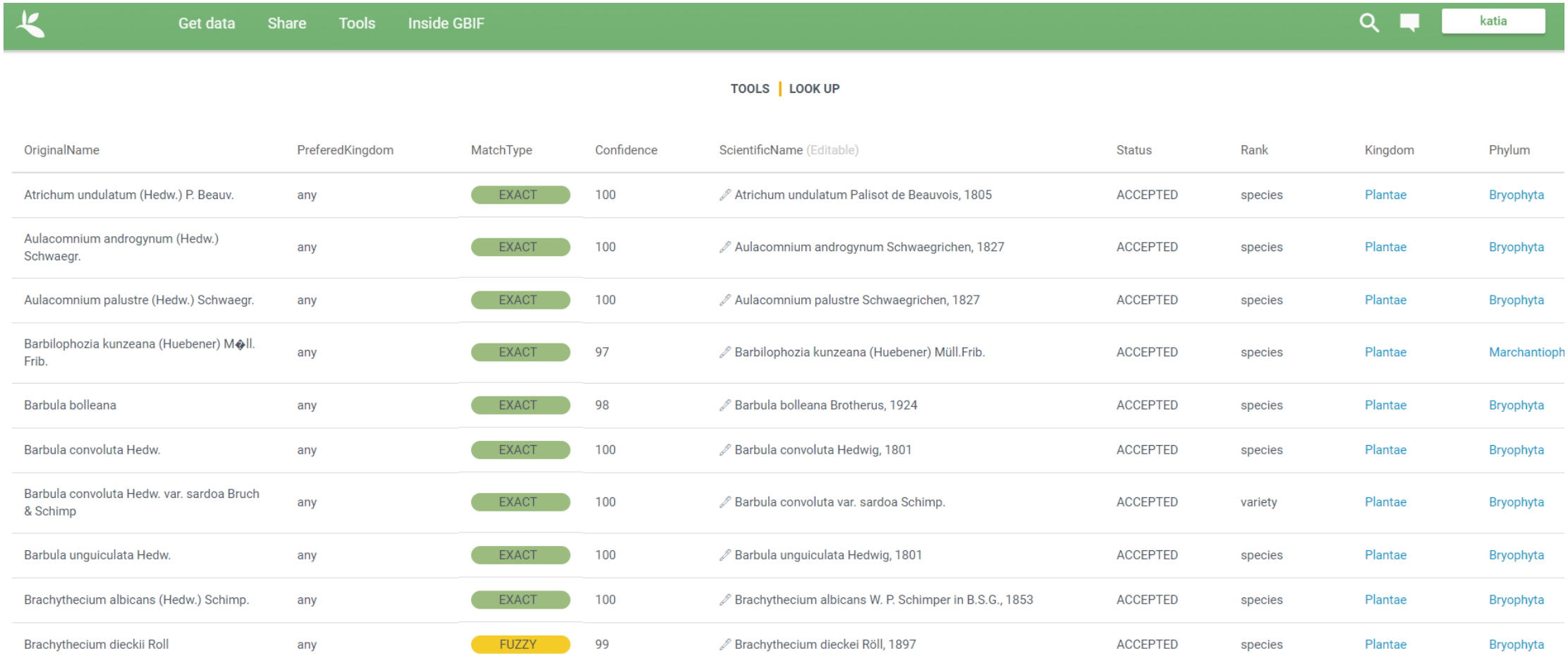
Species matching



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

Tools for managing scientific names

Species matching



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) M◆ll. 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

Tools for managing dates

Tools for managing dates

They allow to parse and transform different date formats.

Date parsing and merging

Canadensys VS date parsing

13-VI-1980	→	1980-06-13
13 June 1980	→	1980-06-13
13-06-1980	→	1980-06-13

Gbif.es

Canadensys

Tools for managing dates

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

Gbif.es

Tools for managing geographical data

Tools related to the management of geographical data

They allow the visualization of locations on a map, the detection of mistakes in coordinates and the conversion of coordinates to the required format (decimal degrees).

Checking

Info XY (species link tools)

Excel to kml (Earth point)

Google Earth, Google Maps,

Carto

Geographical information systems

Conversion

Canadensys coordinate conversion

GBIF.ES coordinate conversion

Geotrans

Gbif.es

speciesLink
data & tools
infoXY

Tools related to the management of geographical data

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

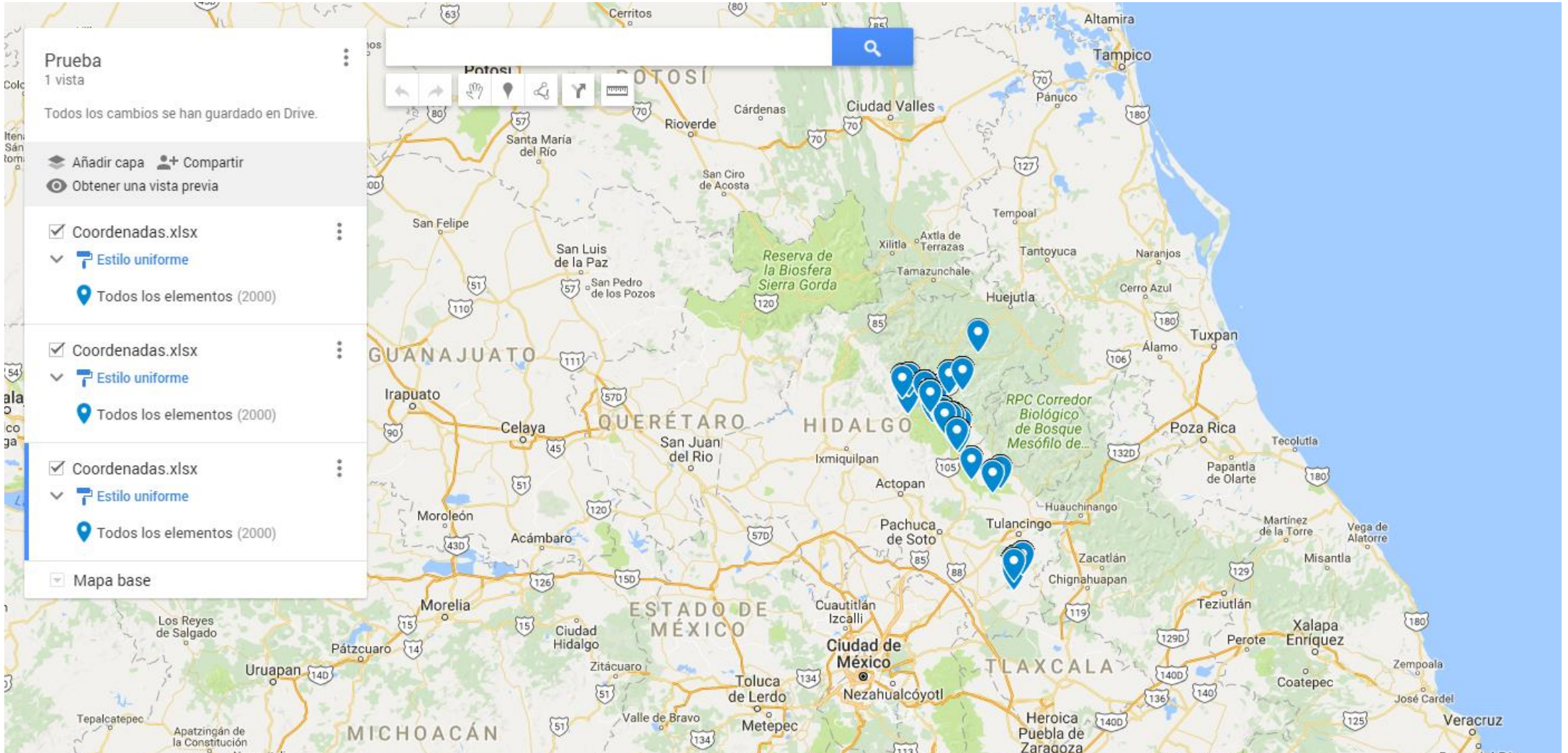
Search

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

Tools related to the management of geographical data



Canadensys

Tools related to the management of geographical data

Coordinate conversion

Use this tool to convert geographic coordinates from DDMMSS 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

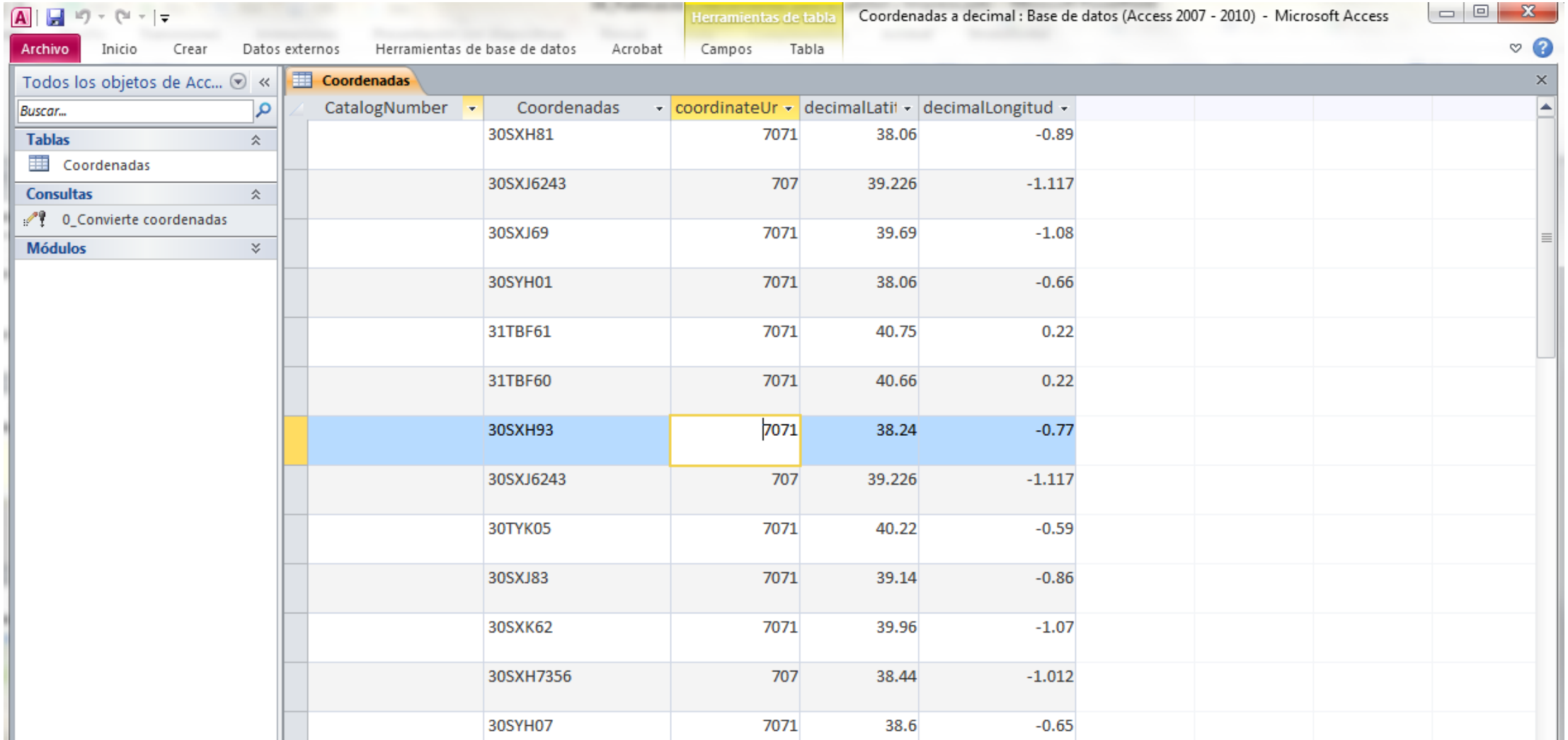
45° 32' 25" N, 129° 40' 31" W

1 | 45.5° N, 129.6° W
2 | 40°26'47"N,74° 0' 21.5022"W

feedback

Gbif.es

Tools related to the management of geographical data



Tools for data validation before publishing

Tools for data depuration and validation before publishing

Specific tools to check and validate that data are correctly standardized and with a sufficient level of quality.

Darwin Test

DATA VALIDATOR

Gbif.es

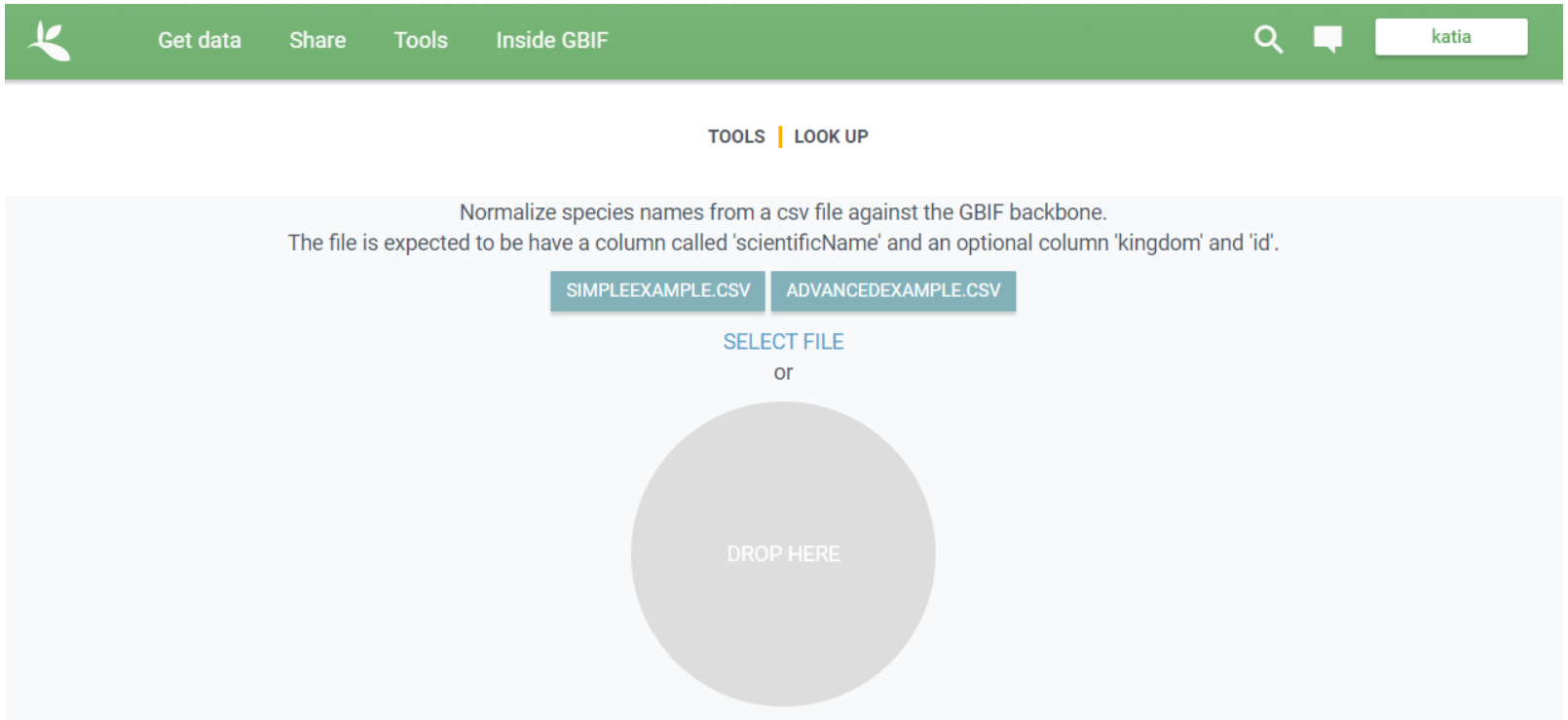
Demos & use cases

Tools for managing scientific names

Species matching

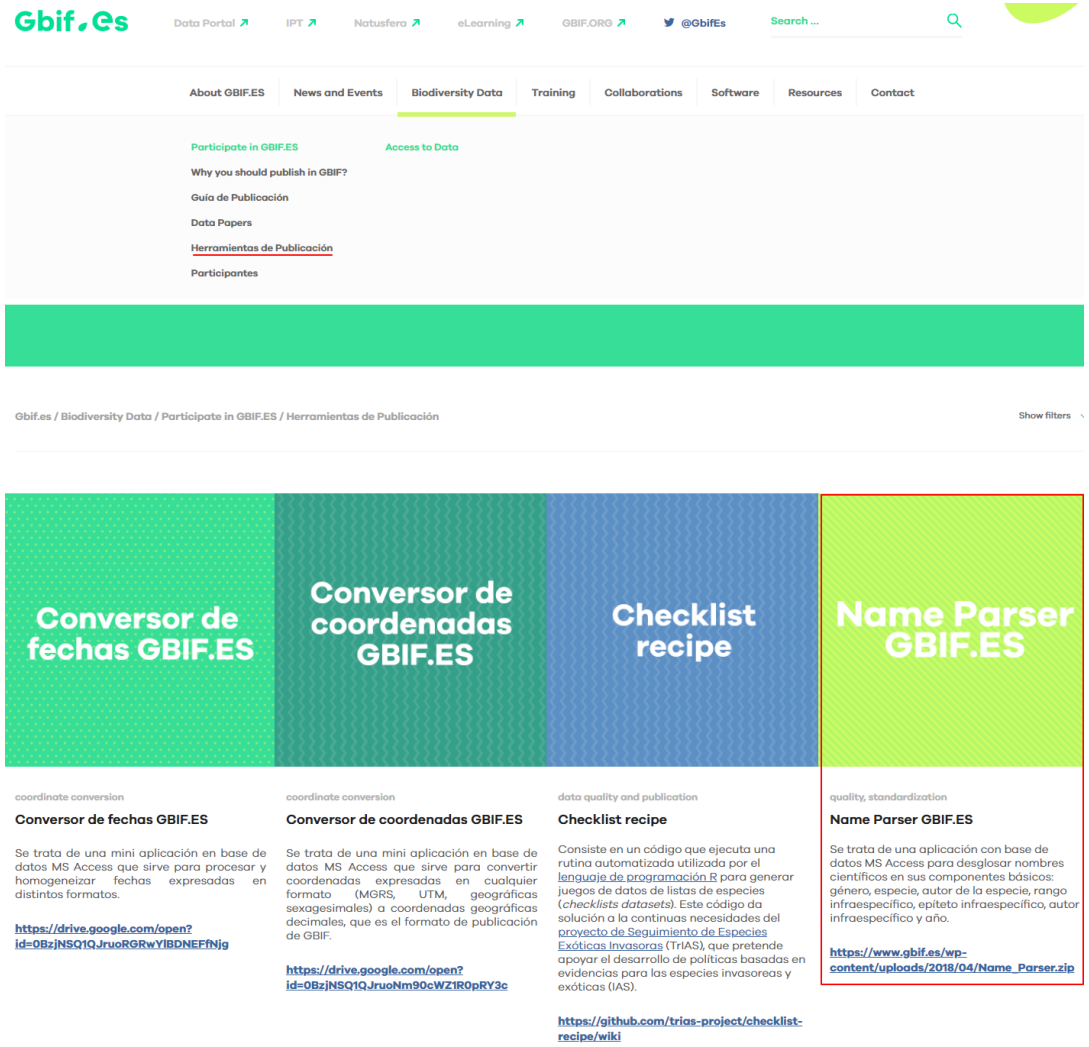
GBIF.org:

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



Gbif.es

Tools for managing scientific names



NAME parser

GBIF Spain:

https://bit.ly/2mrtJXD

Gbif.es

GBIF.ES Name Parser

This tool parses and normalizes all scientific names in their basic components (genus, specific epithet, author, infraspecific range, epithet and author, and year).

1. Download and decompress the GBIF.ES Name Parser tool.
2. Copy the field *ScientificName* from the standard file in the field *name* from the query *Name Parsing* from the Name Parser tool.
3. Apply the query *Name Parsing*.
4. Copy the parsed names (*gen*, *is_sp* and *esspaut*) in the standard file (*specificEpithet*, *infraspecificEpithet* and *scientificNameAuthorship*)

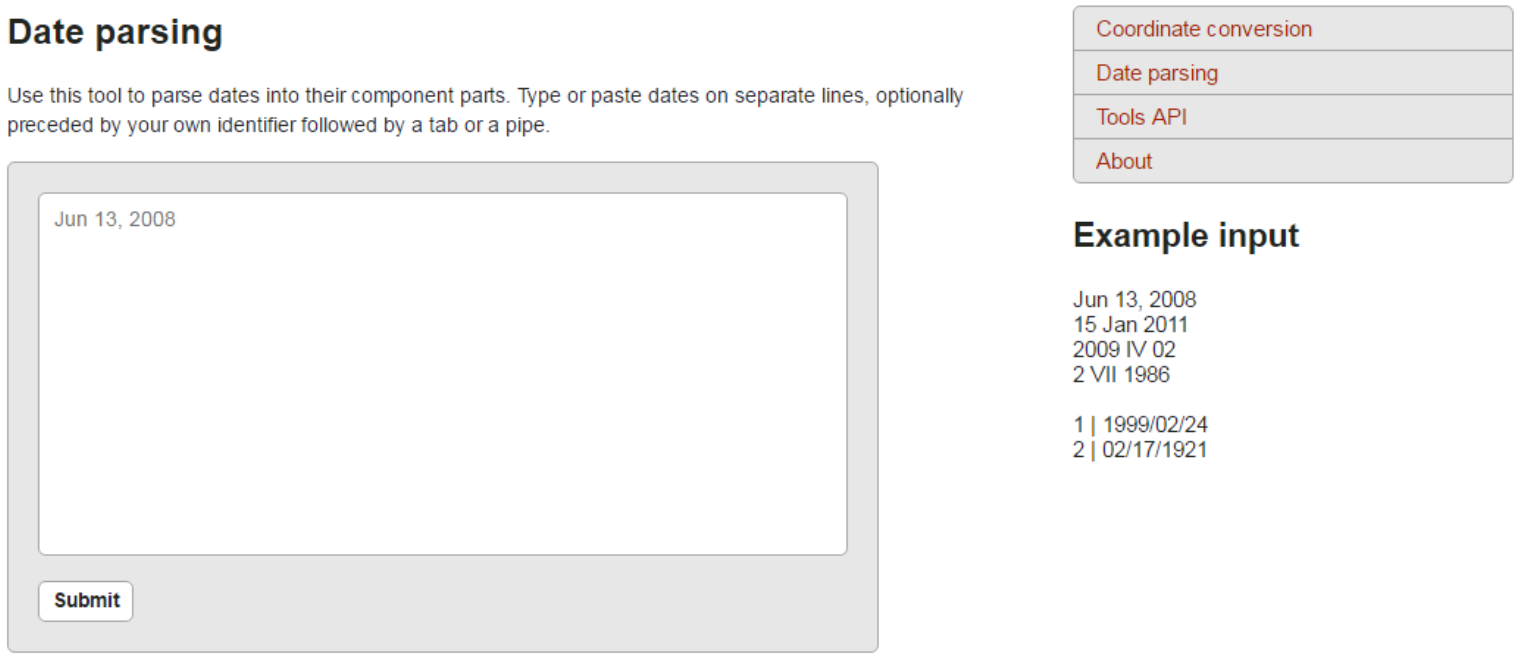
→ USE CASE 3

Gbif.es

Tools for managing dates

Canadensys:

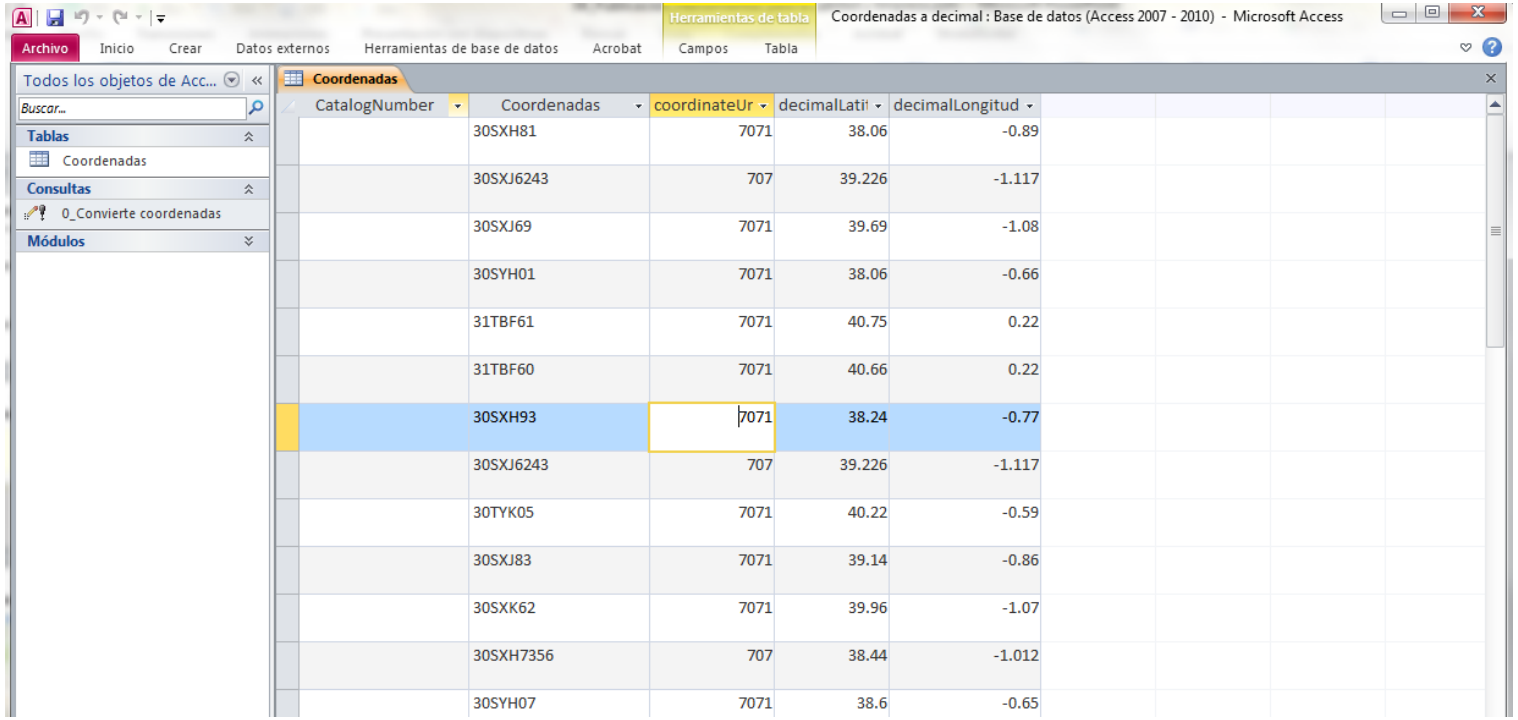
http://data.canadensys.net/tools/dates



Tools for managing geographical data

GBIF Spain:

https://bit.ly/2m5oo7X



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
	30SXK62	7071	39.96	-1.07
	30SXH7356	707	38.44	-1.012
	30SYH07	7071	38.6	-0.65

Gbif.es

Thanks for your attention!

miguel.vega@rjb.csic.es

GBIF Zimbabwe

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

CSIC Consejo Superior de Investigaciones Científicas

REAL JARDÍN BOTÁNICO

Gbif.es

GBIF