Herramientas útiles para la calidad y limpieza de datos

Regional capacity enhancement to Latin America by establishing Chile's node

Katia Cezón
katia@gbif.es
GBIF.ES

CADENA DE INFORMACIÓN DE LOS DATOS DE BIODIVERSIDAD

1 **PLANIFICACIÓN**

No existe

2 **RECOLECCIÓN Y DOCUMENTACIÓN**

- Información taxonómica
- Información espacial
- Datos asociados a la colecta
- Datos descriptivos

3 **DIGITALIZACIÓN**

- Experiencia del personal
- Interpretación de los datos
- Diseño base de datos
- Copias de seguridad

Coste corrección errores

CADENA DE INFORMACIÓN DE LOS DATOS DE BIODIVERSIDAD

Occurrence records	Datasets	Publishing institutions
874,626,060	36,969	1,125

4 CONTROL DE CALIDAD

Feedback

5 PUBLICACIÓN EN GBIF

- Exportación
- Conversiones
- Adaptación de los datos

Coste corrección errores

ASPECTOS A TENER EN CUENTA

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


NO EXISTE
UNA
RECETA

HERRAMIENTAS ÚTILES PARA LA CALIDAD Y LIMPIEZA DE DATOS

Herramientas de 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

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.

OpenOffice.org™

Microsoft® Access

SQLite

Microsoft® Excel

HERRAMIENTAS PARA EL TRATAMIENTO DE 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.

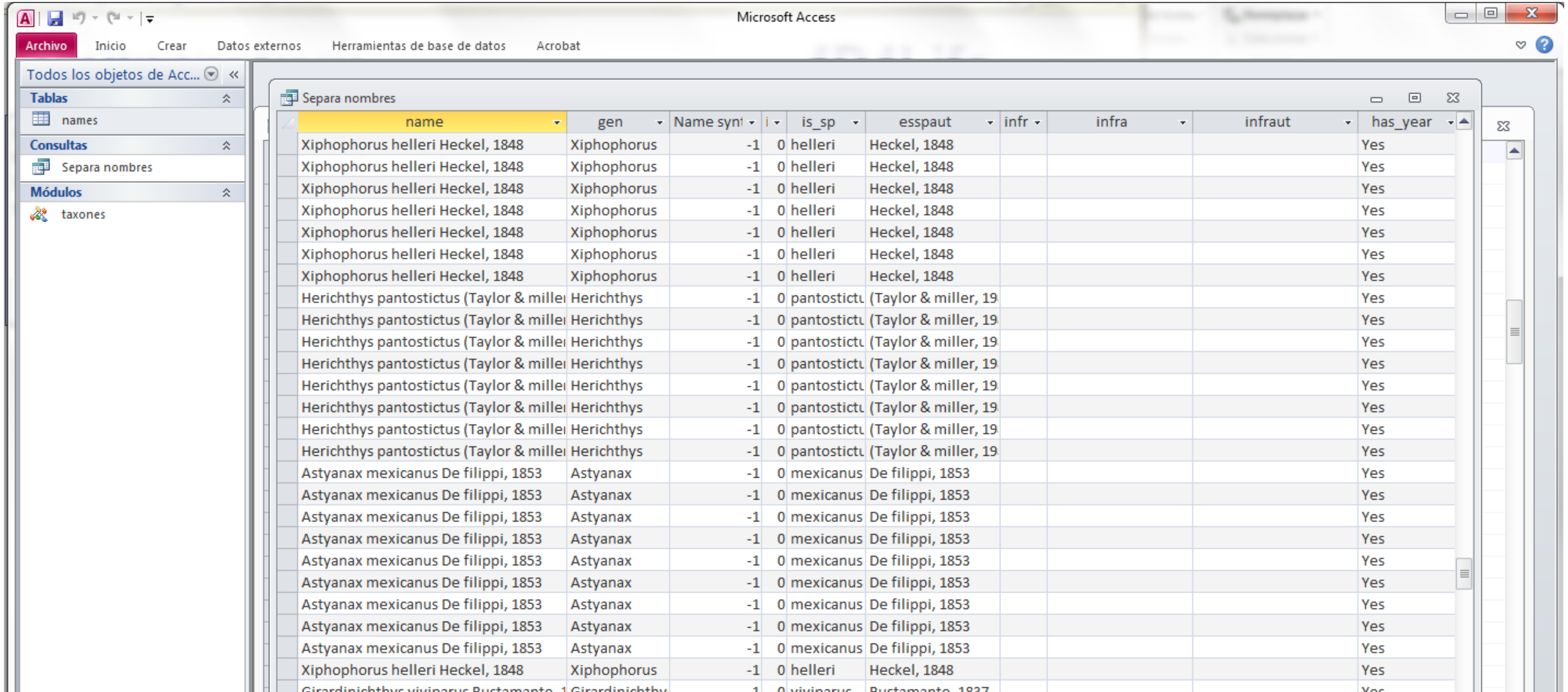
Atomización

- 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 NOMBRES CIENTÍFICOS



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

HERRAMIENTAS PARA EL TRATAMIENTO DE NOMBRES CIENTÍFICOS

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

HERRAMIENTAS PARA EL TRATAMIENTO DE NOMBRES CIENTÍFICOS

T-Rex | Resolución de nombres taxonómicos

Resultados de la busqueda taxonómica

Descargar XLSX

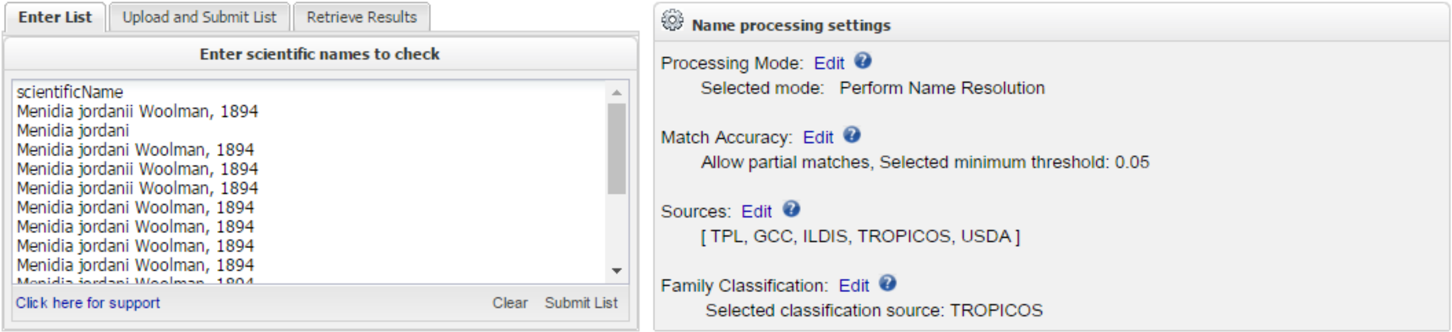
Fuentes seleccionadas: GBIF Backbone Taxonomy

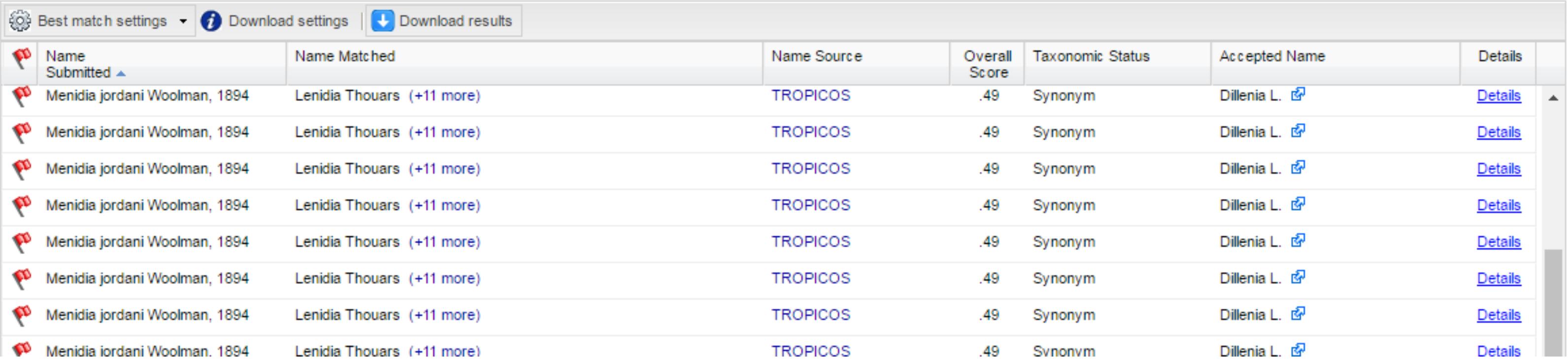
Total Datos 19

Pagina Actual 1 de 2 Anterior Siguiente

Nombre ingresado	Nombre científico resuelto	Fuente de infomación	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

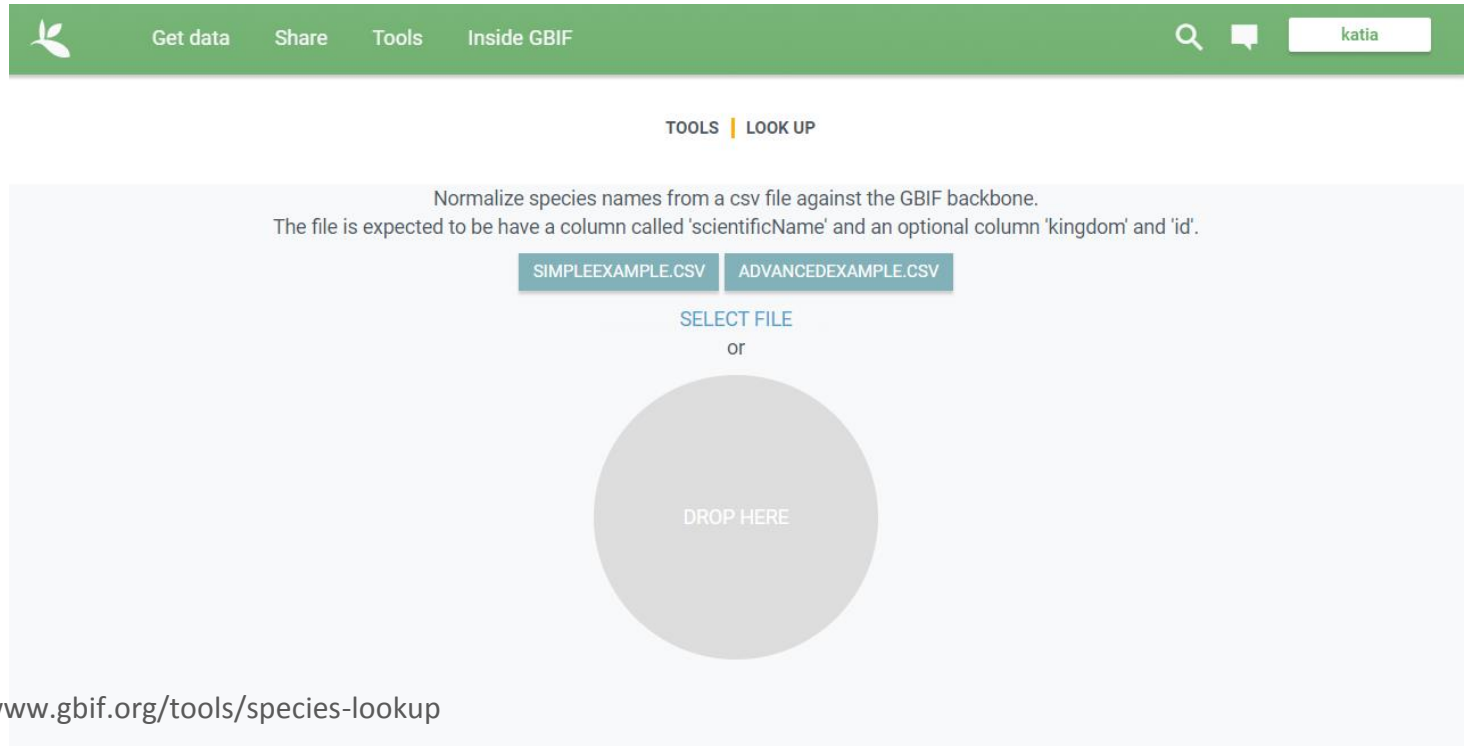
HERRAMIENTAS PARA EL TRATAMIENTO DE NOMBRES CIENTÍFICOS





HERRAMIENTAS PARA EL TRATAMIENTO DE NOMBRES CIENTÍFICOS

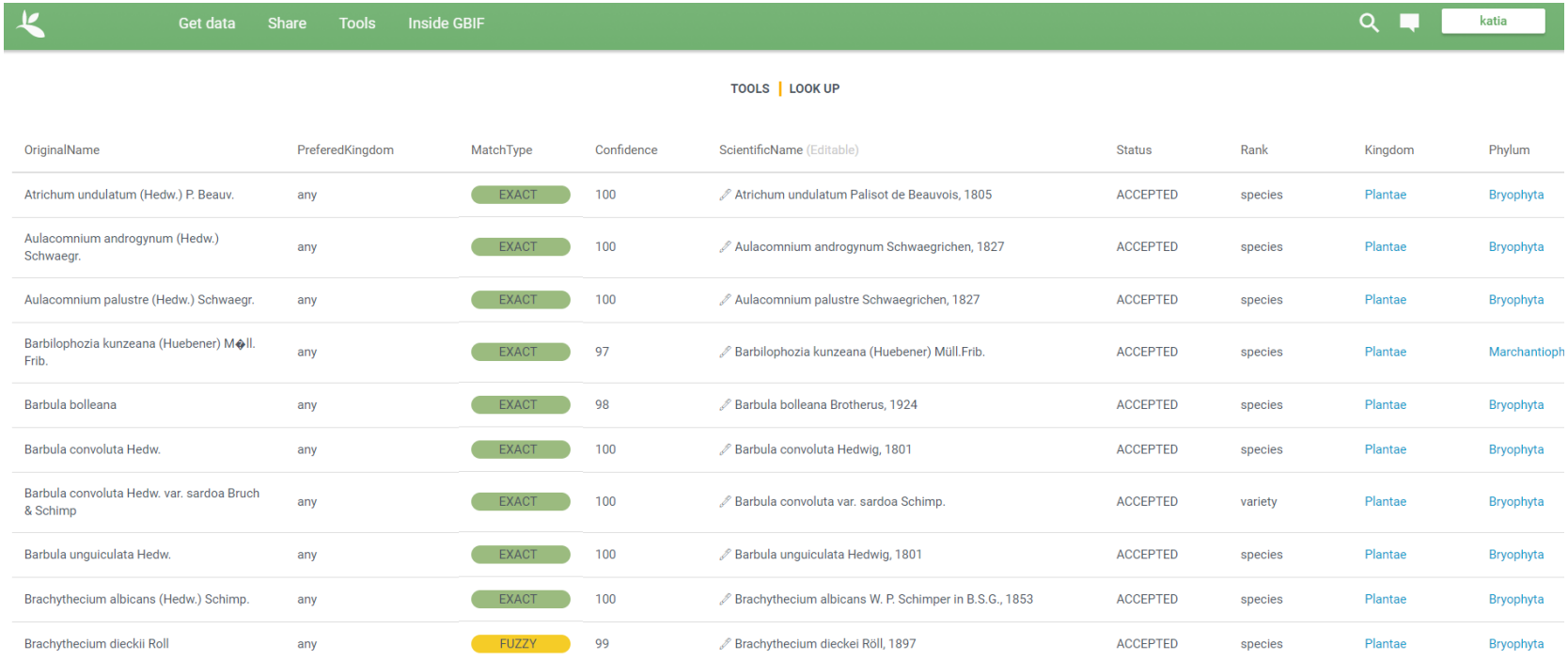
Species matching



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

HERRAMIENTAS PARA EL TRATAMIENTO DE NOMBRES CIENTÍFICOS

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&#65533;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

HERRAMIENTAS PARA EL TRATAMIENTO DE 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

speciesLink
data & tools
infoXY

HERRAMIENTAS PARA EL TRATAMIENTO DE 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

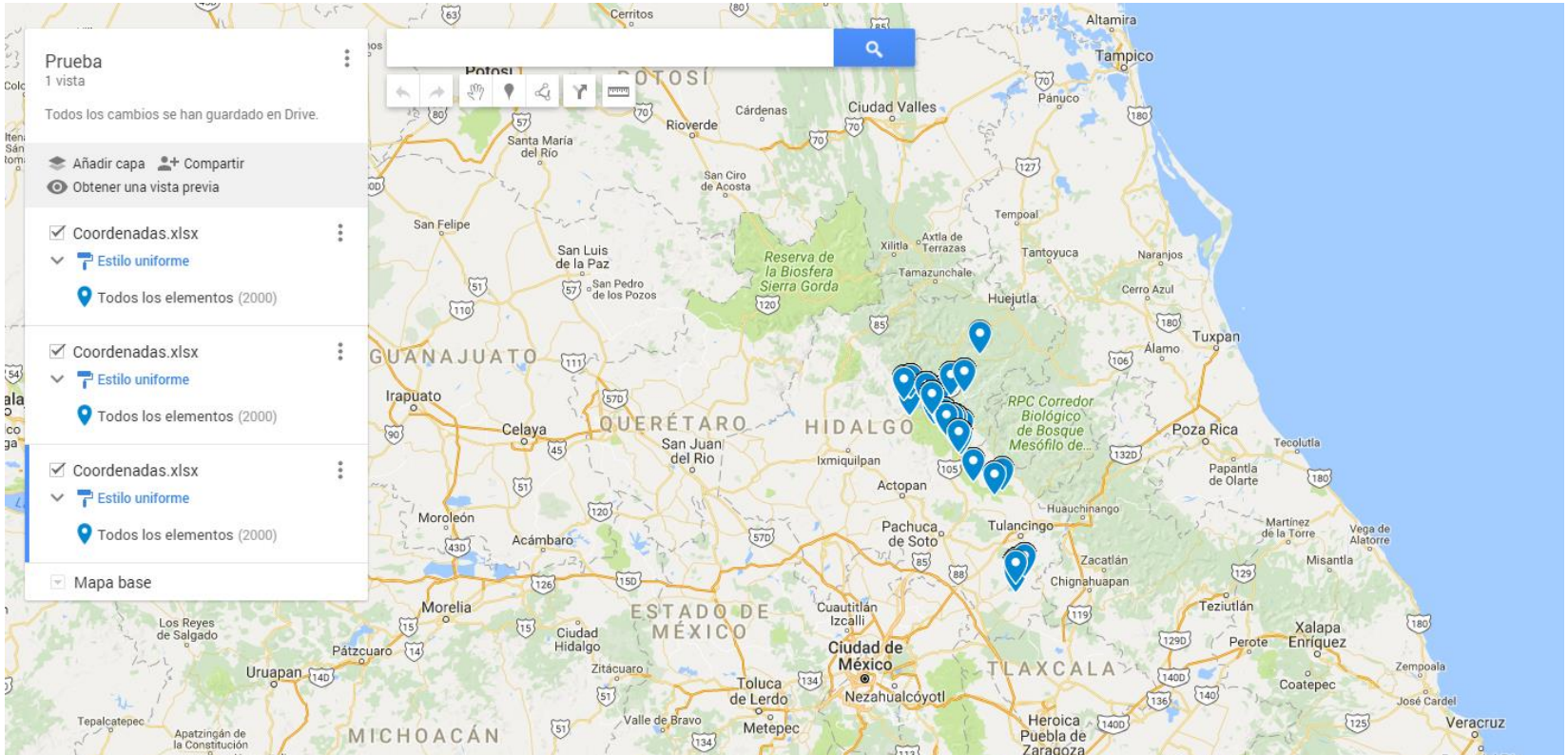
Search

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 DATOS GEOGRÁFICOS



Canadensys

HERRAMIENTAS PARA EL TRATAMIENTO DE DATOS GEOGRÁFICOS

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

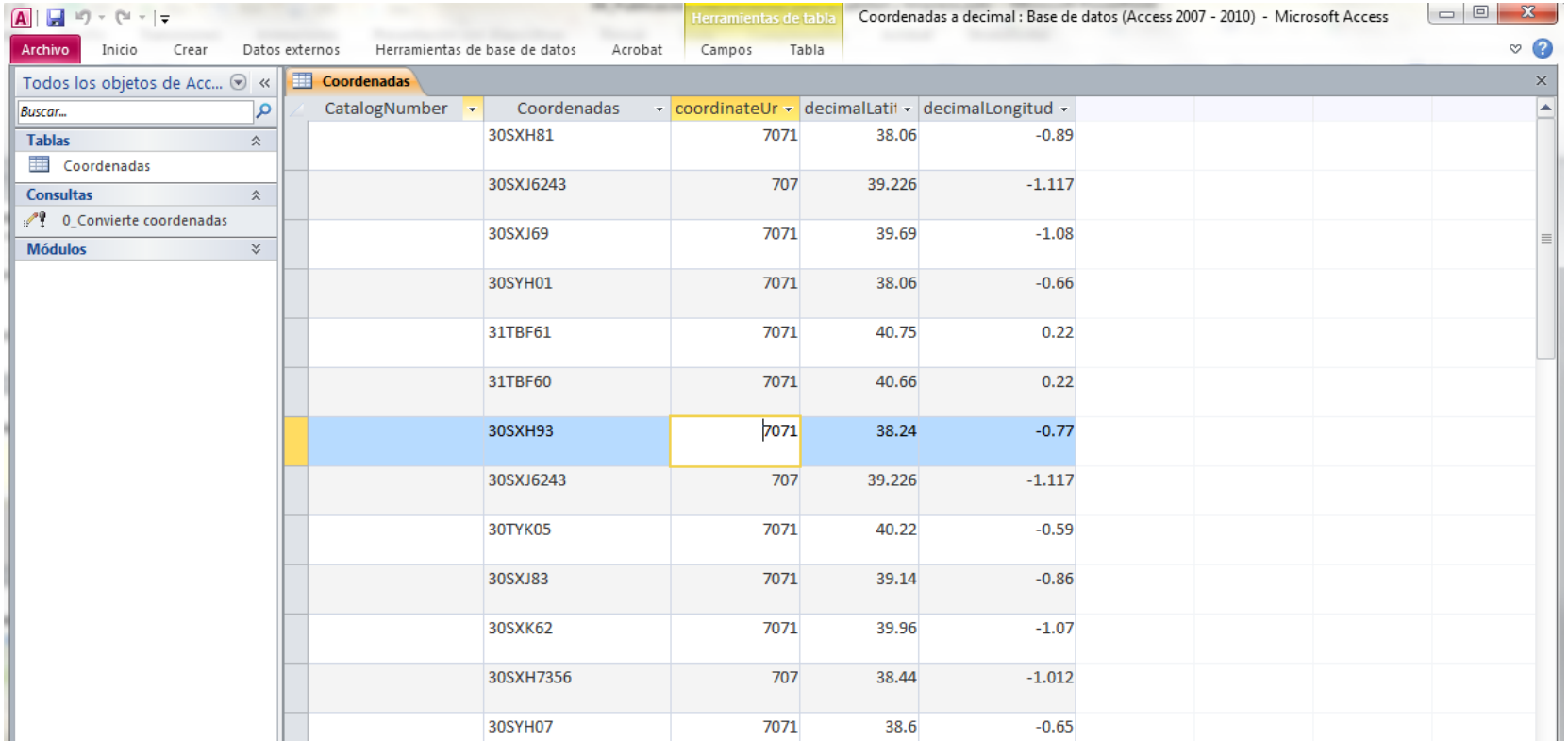
Example input

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

HERRAMIENTAS PARA EL TRATAMIENTO DE DATOS GEOGRÁFICOS



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

HERRAMIENTAS PARA EL TRATAMIENTO DE LAS FECHAS

Herramientas que permiten la unificación y transformación de distintos formatos 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

Canadensys

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

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

HERRAMIENTAS DE VALIDACIÓN Y DEPURACIÓN

OpenRefine

Darwin Test

GBIF Darwin Core Archive Validator

HERRAMIENTAS DE VALIDACIÓN Y DEPURACIÓN

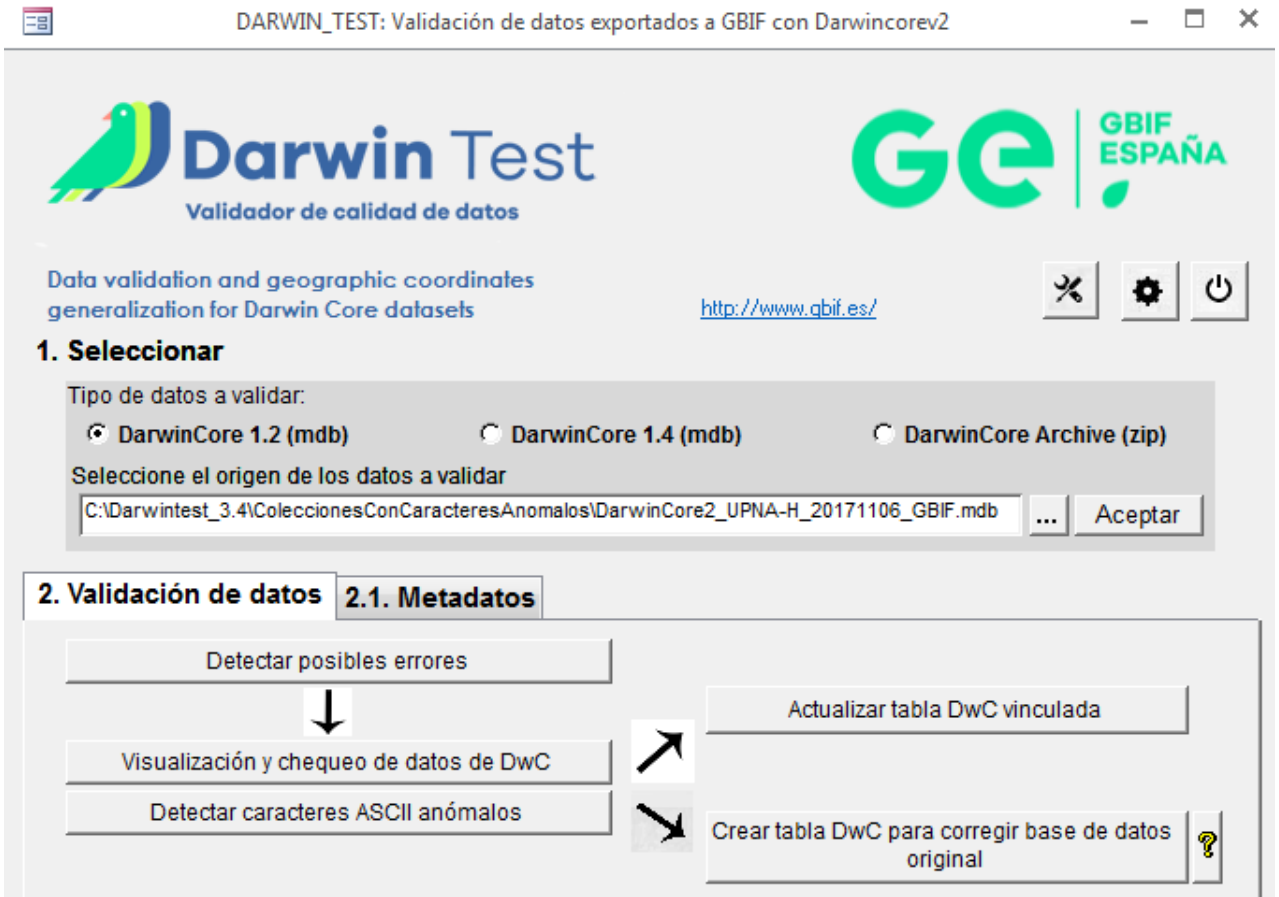
¿Qué es?

Una herramienta de manipulación de datos

Faceting
Clustering
Reconciling

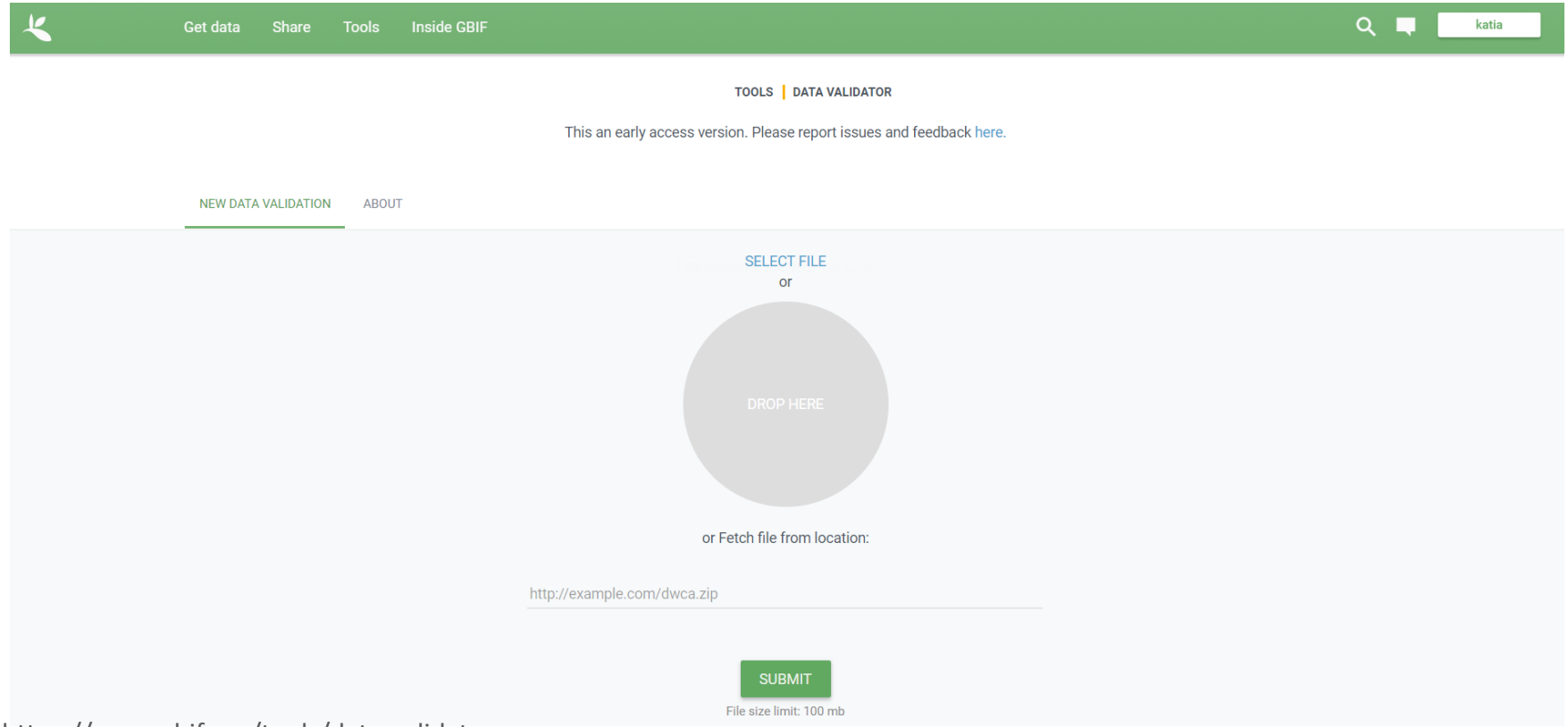
HERRAMIENTAS DE VALIDACIÓN Y DEPURACIÓN

Darwin Test



HERRAMIENTAS DE VALIDACIÓN Y DEPURACIÓN

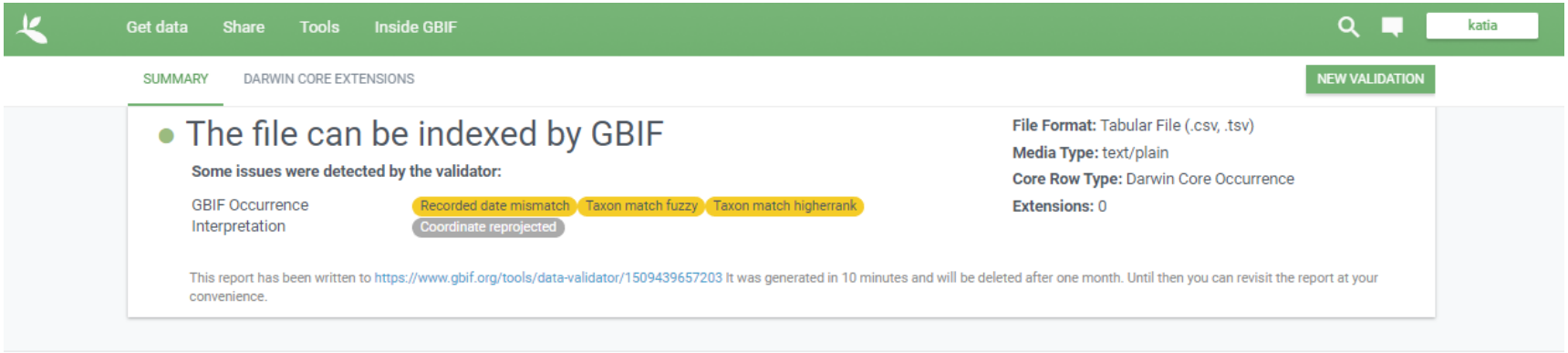
Data validator

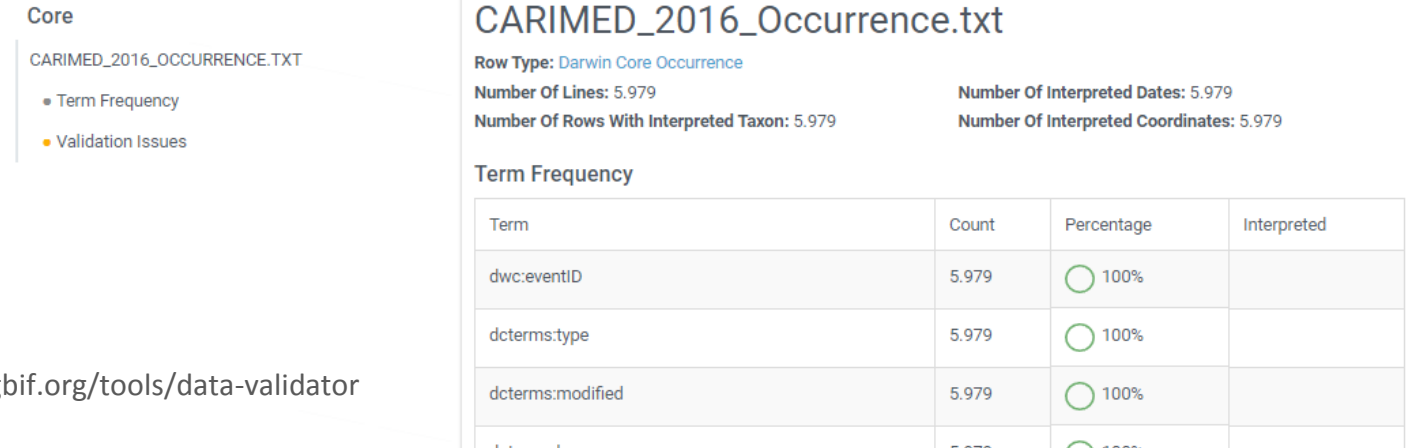


https://www.gbif.org/tools/data-validator

HERRAMIENTAS DE VALIDACIÓN Y DEPURACIÓN

Data validator





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

Regional capacity enhancement to Latin America by establishing Chile's node

Gracias por la atención

UNIDAD DE COORDINACIÓN GBIF.ES
REAL JARDÍN BOTÁNICO-CSIC
katia@gbif.es