Data Use Club

# **Data standards & Processing**

**&** *Practical sessions* 





#### **GBIF Data life cycle**





#### **GBIF Data life cycle – Part 1**



#### Standardisation

Access



### How is data shared on GBIF.org?

Data publishers own and maintain their data, and share it with GBIF





#### Data publishing



The Integrated Publishing Toolkit (IPT): https://www.gbif.org/ipt

Installation



#### Data shared on GBIF is heterogeneous

- Data publishers on GBIF are <u>community-endorsed</u> organizations such as universities, museums, governmental agencies, the private sector, NGOs etc.
- Datasets shared with GBIF come from research projects, national & regional monitoring frameworks, citizen science portals, DNA repositories, collections etc.







#### What is GBIF?



GBIF-the Global Biodiversity Information Facility-is an international network and data infrastructure funded by the world's

🗾 AutoSave Off 🗄 9 × 0 × 📼 Book1 - Ex	ccel 🕐 No Label	ှာ Search					<u> </u>	a x
File Home Insert Page Layout Formulas Dat	ta Review View Automate Developer Help Acrobat	Table Design Query					Comments	台 Share
K19 $\checkmark$ : $\times \checkmark f_x \checkmark$								
A B	c	D	E F	G H		ĸ		
1 gbifID 💌 datasetKey	occurrenceID	× kingdom	phylum 💌 class 💌	order Tamily ge	enus <b>v</b> species	infraspecificEpithet taxonRan	k 🕶 scientificNa	ame
2 3784774669 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/116947197	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	thocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3 3747328581 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/108897781	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4 3802272692 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/117905444	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
5 3109046931 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/75180064	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
6 3112906594 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/76668819	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
7 3070346935 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/43314305	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
8 3337601527 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/90426259	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
9 2611243059 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/43734042	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamine
0 2619886267 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/44737037	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
1 4014949558 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/121363054	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
2 3947748256 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/138594492	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3759762318 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/111795095	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4 3090801655 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/44854742	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
5 3772504001 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/115102066	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
6 3985635046 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/142132874	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
7 3118407520 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/79365628	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
18 2597708992 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/41292060	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
19 3112338632 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/75823175	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	SPECIES	Anthocharis	cardamin
4908254275 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/229029951	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	S SPECIES	Anthocharis	cardamin
21 3764126203 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/113113154	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
22 3302330333 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/82455948	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3759477609 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/112127149	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4591955553 12d04ddc-d939-405b-81be-807ebca3ed18	563062e1-8cba-11eb-8ba1-00505688c2ac	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4591963599 12d04ddc-d939-405b-81be-807ebca3ed18	cd3f93be-ff6c-407f-a9c3-997adca09308	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4150194584 57803fb7-d11d-4d81-b6b9-0bd8d9de57f6	76c81ceb-9bf2-41ed-ab47-020000083133	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4150195177 57803fb7-d11d-4d81-b6b9-0bd8d9de57f6	76aec9d7-9bf2-41ed-ab47-02000083133	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
28 2620151939 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/44780261	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3859984623 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/73927684	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
5007902115 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/112681239	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3760152131 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/111620522	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
22 2883188438 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/62030477	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
33 3097161130 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/74510260	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
34 3772540189 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/115127162	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
35 5007544058 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/256509775	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
36 3764106850 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/112736276	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
3118335042 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/79027300	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamine
3858359707 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/119434915	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	thocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
2626422128 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/46081885	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	thocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
40 4015259534 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/146292445	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
41 3764209996 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/112951775	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	S SPECIES	Anthocharis	cardamin
2 3772675298 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/114580874	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	thocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
43 3858534776 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/120660274	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	hthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
4421235924 50C9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/185552323	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	ntnocharis Anthocharis cardamine	S SPECIES	Anthocharis	cardamin
20111934/5 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/44389997	Animalia	Arthropoda Insecta	Lepidoptera Pierídae Ar	nthocharis Anthocharis cardamine	S SPECIES	Anthocharis	cardamin
40 37/3334211 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/115861125	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	S SPECIES	Anthocharis	cardamini
7 3//318/489 50C9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/116054627	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin
8 4172286655 50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/177382240	Animalia	Arthropoda Insecta	Lepidoptera Pieridae Ar	nthocharis Anthocharis cardamine	s SPECIES	Anthocharis	cardamin



### To which extent is there content in the fields you download?



- gbifID: 100%
- habitat: 7.1%
- individualCount: 62.4%\*
- stateProvince: 80.8%\*
- scientificName: 98.7%
- depth: 2.4%
- elevation: 5.6%
- lifeStage: 9.1%
- year: 95.9%





#### Data standards

Searchable fields on GBIF.org are provided and maintained by a global biodiversity community



### Data provided to GBIF is usually shared in a Darwin Core Archive (DwC-A)

Other sharing options exist, but a GBIF user will always download a DwC-A

- Darwin Core is a data standard managed by Biodiversity Information Standards (TDWG)
- Data standards *define the field name* and sometimes the *value* content of the field

For example, a data publisher want to might share columns called:



https://www.tdwg.org/standards/dwc/ | https://www.tdwg.org/ | https://www.gbif.org/ipt





### The fields available for sharing depends on the dataset class

Datasets shared with GBIF are shared as either:

- Metadata-only: datasets describing unavailable resources like undigitized natural history collections or data that are not yet standardized
- Checklists: a catalogue or list of named organisms, or taxa
- Occurrence: the evidence of the occurrence of a species (or other taxon) at a particular place on a specified date
- Sampling events: offering evidence that a species occurred at a given location and date, but also making it possible to assess community composition for broader taxonomic groups or even the **abundance of species** at multiple times and places







Occurrences

### Core dataset\*

\* taxon, occurrence or event





### **Missing fields relevant for your analysis?**

- You can propose new terms or changes to terms for Darwin Core or its associated extensions
- Adding or changing data standards is a community-driven process which requires engagement
- Please consider:
  - Do the term already exist, but would require an update?
  - Where would a new term best be placed? Which Class (taxon, event, occurrence, ۲ location etc)?
  - Is the term better suited in an extension or already present in an extension?
  - What would the addition or change facilitate?

Propose new terms or changes in 'issues': <u>https://github.com/tdwg/dwc</u>







### Dataset metadata is described in Ecological Metadata Language

- Publishers can share information about the dataset origins with GBIF's EML Profile, for example:
  - People involved in generating and maintaining the data
  - The overall spatial & temporal scope of the data
  - The taxonomic scope of the dataset
  - Funding & project information
  - Methods & design

The current GBIF EML Profile: https://rs.gbif.org/schema/eml-gbif-profile/1.3/eml-gbif-profile.xsd







#### Access



### **GBIF** Data Processing

# 1. Normalization

# 2. Flagging

## 3. Enrichment



### Typical GBIF normalization doesn't change much

Get data

How-to

Tools

Community About

OCCURRENCE 20 JULY 2014

### Calopteryx splendens (Harris, 1780)

Banded Demoiselle In English Observed in France

Animalia > Arthropoda > Insecta > Odonata > Calopterygidae > Calopteryx

DETAILS

Species: Calopteryx splendens (Harris, 1780) Location: Europe > France Basis of record: Human observation



-/- 🛪 🔍 🗖

jwaller

Synced 2 years ago

3

Event ID: def6f3e4-5a14-43d4-8bae-2fc34e1eec7f Dataset: Liste rouge des Odonates de France-Liste rouge des Odonates de métropole 2016 Publisher: UMS PatriNat (OFB-CNRS-MNHN), Paris

Issues: Continent derived from coordinates







### Basis of record normalization

🤾 Get data Hov	v-to Tools C	ommunity About			-/- x <sub>A</sub> Q	jwaller
< Occurrences Simple filters All file	inters	12	SEARCH OCCURRENCES	3,113,776,236 RESULTS		
Occurrence status	~	TABLE GALLERY MAP TAXONOMY	METRICS & DOWNLOA	AD		
Licence	~	Scientific name	Country or area	Coordinates	Event date	Occurrence
Scientific name	~	Convolvulus angustissimus subsp. angustis	Australia	35.4S, 149.1E	2025 Jan 08	Present
Basis of record	^	Pheidole Westwood, 1839	Australia	35.3S, 149.1E	2025 Jan 10	Present
<ul> <li>Observation</li> <li>Machine observation</li> </ul>	16,764,229 24,147,137	Hypericum perforatum L.	Australia	33.7S, 148.3E	2025 Jan 03	Present
<ul> <li>Human observation</li> <li>Material sample</li> </ul>	2,710,359,619 65,796,690	<ul> <li>Stigmatium victoriae</li> </ul>	Australia	35.4S, 149.0E	2025 Jan 14	Present
Material citation  Preserved specimen	8,189,400 256,167,603	Pavo cristatus Linnaeus, 1758	Australia	35.3S, 149.1E	2025 Jan 12	Present
Fossil specimen	9,648,620	Turdus merula Linnaeus, 1758	Denmark	57.0N, 9.9E	2025 Jan 03	Present
<ul> <li>Living specimen</li> <li>Occurrence</li> </ul>	2,032,823 20,670,115	Ornithorhynchus anatinus (Shaw, 1799)	Australia	42.9S, 147.3E	2025 Jan 02	Present
Year	~	Daviesia latifolia R.Br.	Australia	34.6S, 149.8E	2025 Jan 16	Present
Month	~ -	Urtica dioica L.	Denmark	55.8N, 12.5E	2025 Jan 02	Present





### Normalization of **Basis of Record**

GBIF processing is focused on **normalizing** data, which is taking variable publisher supplied values and standardizing them to a controlled vocabulary.

103 unique values given by publishers for "Preserved Specimens"

Publisher supplied value **Pinned Insect** 



Publisher supplied value **Dried Specimen** 



Publisher supplied value

#### Papered specimen



Publisher supplied value

#### Museum specimen



GBIF interpreted value **Preserved specimen** 



Normalization using Vocabulary Server

Most data normalization occurs with **hard coded** mappings in GBIF processing code. However, GBIF is slowly moving to a more flexible system called the **vocabulary server**.

Only a few terms currently use the **vocabulary server** for normalization, but more are planned:

- ✓ EstablishmentMeans
- ✓ LifeStage
- ✓ OccurrenceStatus
- ✓ DegreeOfEstablishment
- ✓ Pathway



Normalization of **lifestage** with the **vocabulary server** 

The vocabulary server will allow for more accurate and transparent mapping of publisher supplied values to **controlled vocabularies**.

**4,276** unique values for **lifestage** are now mapped to a controlled vocabulary of **40** terms.

Publisher supplied lifestage	GBIF interpreted
70 adultos	Adult
250300 ad	Adult
HI2. Infantil 2.	Juvenile
16 egg	Egg
fledged young	Juvenile
11J	Juvenile
62 Subadult(s)	Subadult
Larva/nymph	Nymph



### Lifestage normalization

Get data How-to	Tools Co	ommunity About			-^- ≭ <sub>A</sub> Q ■	jwaller
< Occurrences			SEARCH OCCURRENCES	3,113,776,236 RESULTS		
Organism quantity	~		METRICS 👱 DOWNLO	AD		
Organism quantity type Relative organism quantity	~	Scientific name	Country or area	Coordinates	Event date	Occurrenc
Sex	~	Convolvulus angustissimus subsp. angustis	Australia	35.4S, 149.1E	2025 Jan 08	Present
Life stage	~	Pheidole Westwood, 1839	Australia	35.3S, 149.1E	2025 Jan 10	Present
Search		Hypericum perforatum L.	Australia	33.7S, 148.3E	2025 Jan 03	Present
Establishment means	~	<ul> <li>Stigmatium victoriae</li> </ul>	Australia	35.4S, 149.0E	2025 Jan 14	Present
Degree of establishment	~	Pavo cristatus Linnaeus, 1758	Australia	35.3S, 149.1E	2025 Jan 12	Present
Pathway	~	Turdus merula Linnaeus, 1758	Denmark	57.0N, 9.9E	2025 Jan 03	Present
Media type	~	Ornithorhynchus anatinus (Shaw, 1799)	Australia	42.9S, 147.3E	2025 Jan 02	Present
GBIF ID	~	Daviesia latifolia R.Br.	Australia	34.6S, 149.8E	2025 Jan 16	Present
Provenance	-	Urtica dioica L.	Denmark	55.8N, 12.5E	2025 Jan 02	Present





### **GBIF** Taxonomic interpretation

Name with authorship Calopteryx aequabilis Say, 1840



Just the name Calopteryx aequabilis



Mis-spelling Calopteryx aequabile



Incomplete name **Calopteryx sp.** 



Calopteryx aequabilis Say, 1840

Calopteryx aequabilis Say, 1840

Calopteryx aequabilis Say, 1840 taxon match fuzzy

Calopteryx Leach, 1815 taxon match higherrank



### **GBIF** Flagging



**Technical Documentation** 

Data processing Data use API

#### Data Use

- Data downloads
- Issues and flags

#### **Occurrence issues and flags**

- Checklist issues and flags
- **API Usage**
- **API** Downloads
- Cloud computing services

Citation

Data Use / Issues and flags / Occurrence issues and flags 命

# Occurrence issues and flags

More than 60 issues and flags have been created to deal with common data quality problems. Data users may use these flags to exclude occurrences from their analysis, or to apply their own data quality checks.

#### Geospatial issues

Flag name	Definition	Terms	Example
Zero coordinate	Coordinates are exactly 0/0, often indicating an actual null coordinate.	dwc:decimalLatitude, dwc:decimalLongitude	example
Coordinate out of range	The supplied coordinates lie outside of the range for decimal lat/lon values (-90/90, -180/180).	dwc:decimalLatitude, dwc:decimalLongitude, dwc:verbatimCoordinates,	example

rage contents	Pa	ge	contents
---------------	----	----	----------

Geospatial issues Taxonomic issues Date issues Vocabulary issues GRSciColl issues Other issues





### Country Coordinate Mismatch F

K Get data	How-to	Tools	Community	Abo		
< Occurre	ences		2			
Search all fields		Q	TABLE	GALLERY	MAP	TAXONOMY
Simple Ad	lvanced		+			20 00 0
Occurrence status		~	-			
Licence		$\sim$	7			
Scientific name		~	0			
Basis of record		~	P			
Location		~			8 ° 8	
Administrative areas (gadm.or	g)	~				
Coordinate uncertainty in metr	res	~				
Year		~	• * •		8 000	
Month		$\sim$				
Dataset		~	°00			
Country or area		~	~ ~	2 849	8 6 6	090 0



METRICS **±** DOWNLOAD





### Suspicious Coordinates ("has geospatial issue")

"This map contains occurrences flagged by GBIF as having suspicious coordinates."

#### Zero-zero coordinate

This refers to geographic coordinates (latitude, longitude) recorded as (0, 0. This is often used as a placeholder for missing or unknown coordinates.

#### Country coordinate mismatch

This occurs when the recorded geographic coordinates fall outside the country specified in the record.

#### Coordinate invalid

This indicates that the recorded geographic coordinates are not valid (bad input).

#### Coordinate out of range

This refers to coordinates that fall outside world geographic bounds.





### Suspicious records |

K	Get data	How-to	Tools	Community	About		
<	Occur	rences	ļ	1			
Search all	fields		Q	TABLE	GALLERY	MAP	Т
	Simple filters	All filters		+		•••	
Occurrence	e status		~	-	2468	689	
Present	t			7	1845		
Licence			~		6 s 84		1
Scientific n	ame		$\sim$			62	6
Basis of re	cord		~				Ę
Year			~		201	56.7	
Month			~				
Location			~	180		000	ę
Administra	tive areas (gadm.	org)	~	This map	contains occurre	ences flago	ed b
Country or	area		~	© OpenSt	reetMap contribu	utors, © Op	enM

TAXONOMY METRICS





### Inferring values for occurrenceStatus from other fields



Publisher supplied value **Present** 



Publisher supplied value **Absent** 



Publisher supplied value **Null** 



Publisher supplied value **Null** 

Individual count = 0



Publisher supplied value

#### Null

Individual count = 0 Basis of Record = Preserved Specimen



Publisher supplied value

**Absent** Individual count = 3 GBIF inferred **Present** 

GBIF inferred **Absent** 

GBIF inferred **Present** 

**GBIF** inferred

#### Absent

occurrence status inferred from individual count

**GBIF** inferred

#### Present

occurrence status inferred from basis of record

**GBIF** inferred

#### Present

individual count conflicts with occurrence status



### **GBIF** Geocoding (enrichment using coordinates)



### **GADM** (Global Administrative Areas)





### **GBIF** continent



### Somewhat complex **Continent** processing ...

![](_page_32_Picture_1.jpeg)

Publisher supplied value **Europe** 

![](_page_32_Picture_3.jpeg)

Publisher supplied value **Africa** lat-lon in **Europe** 

![](_page_32_Picture_5.jpeg)

Publisher supplied value **Null**lat-lon in **Europe** 

![](_page_32_Picture_7.jpeg)

Publisher supplied value

lat-lon is **Null** Country is **Denmark** 

![](_page_32_Picture_10.jpeg)

Publisher supplied value **Null** 

lat-lon is **Null** Country is **Türkiye** 

- GBIF inferred **Europe**
- GBIF inferred
- Africa
- Continent coordinate mismatch
- GBIF inferred
- Europe
- Continent derived from coordinates
- **GBIF** inferred
- Europe
- Continent derived from country
- GBIF inferred

![](_page_32_Picture_24.jpeg)

### **GBIF** continent

#### dwc:continent processing is somewhat complex than most terms.

Get data How-to	o Tools	Community	About
Occurrences	-		
	~		
h	~	TABLE	GALLERY
ation	~	+	
ncluding coordinates		- B	ulgaria
ministrative areas (gadm.org)	$\sim$	7	
untry or area	~		QE
Türkiye			-
itinent	~		- 3 <b>2</b>
Europe			
ataset	~		
ublisher	~		
JCN Global Red List Category	~	This map of	contains occu

https://www.gbif.org/occurrence/map?continent=EUROPE&country=TR&has\_coordinate=true&has\_geospatial\_issue=false&occurrence\_status=present

![](_page_33_Figure_5.jpeg)

![](_page_33_Picture_6.jpeg)

### Filtering centroids

K	Get data	How-to	Tools	Со	mmunity	About		
<	Occur	rences	Ţ	2			$\square$	
Identified by	у		~	*				
Identified by	y ID		~		TABLE	GALLERY	MAP	TAXON
	Oth	er			+			
Issues and f	flags		~		<b>K</b>		175	
Distance fro	om centroid in me	etres	~		0			
🔽 Is 0								8.g
Repatriated	U.		~					
Has related	records		~					•
DWCA exter	nsion		~					÷.
Last interpr	eted		~					Ę
					This map co	ontains occurre	ences flagg	ed by GBII
				-	© OpenStre	etMap contrib	utors, © Op	enMapTile

![](_page_34_Figure_4.jpeg)

![](_page_34_Figure_5.jpeg)

![](_page_34_Picture_7.jpeg)

### Country/province/area centroid source

![](_page_35_Picture_1.jpeg)

https://github.com/jhnwllr/catalogue-of-centroids

type 🛑 PCLI 🜑 ADM1

![](_page_35_Picture_4.jpeg)

### **Enriching GBIF data with IUCN threat statuses**

![](_page_36_Figure_1.jpeg)

-/- XA Q U

jwaller

CHECKLIST DATASET REGISTERED JULY 18, 2011

### The IUCN Red List of Threatened Species

Published by International Union for Conservation of Nature

254,583 RECORDS **4 CITATIONS** 

> INGEST NOW LOGS

![](_page_36_Picture_10.jpeg)

Publication date: May 9, 2023 Metadata last modified: November 17, 2023 Hosted by: GBIF Secretariat Licence: CC BY 4.0

DOI 10 15468/0anb58

![](_page_36_Picture_13.jpeg)

![](_page_36_Picture_14.jpeg)

![](_page_36_Picture_15.jpeg)

### Enriching GBIF data with IUCN threat statuses

🤾 Get data How-	to Tools	Community About		- <b>/-</b> ≭ <sub>A</sub> Q ■	jwaller
< Occurrences			SEARCH OCCURRENCES 3,113,776,236 RESULTS		
Continent	~	TABLE CALLERY MAD TAXONOMY			
Dataset	~		METRICS <u>POWNEOAD</u>		
Publisher	~	Scientific name	Country or area Coordinates	Event date	Occurrence stat
IUCN Global Red List Category	~	Convolvulus angustissimus subsp. angustis	Australia 35.4S, 149.1E	2025 Jan 08	Present
Extinct (EX)	76,032	Pheidole Westwood, 1839	Australia 35.3S, 149.1E	2025 Jan 10	Present
<ul> <li>Extinct in the Wild (EW)</li> <li>Critically Endangered (CR)</li> </ul>	16,441 2.736.116	Hypericum perforatum L.	Australia 33.7S, 148.3E	2025 Jan 03	Present
Endangered (EN)	6,577,196	<ul> <li>Stigmatium victoriae</li> </ul>	Australia 35.4S, 149.0E	2025 Jan 14	Present
<ul> <li>Near Threatened (NT)</li> </ul>	53,282,736	Pavo cristatus Linnaeus, 1758	Australia 35.3S, 149.1E	2025 Jan 12	Present
<ul><li>Least Concern (LC)</li><li>Data Deficient (DD)</li></ul>	2,122,926,035 7,112,915	Turdus merula Linnaeus, 1758	Denmark 57.0N, 9.9E	2025 Jan 03	Present
Not Evaluated (NE)	739,693,434	Ornithorhynchus anatinus (Shaw, 1799)	Australia 42.9S, 147.3E	2025 Jan 02	Present
Issues and flags	~	Daviesia latifolia R.Br.	Australia 34.6S, 149.8E	2025 Jan 16	Present
https://www.gbif.org/occurrence/map		Urtica dioica L.	Denmark 55.8N, 12.5E	2025 Jan 02	Present

![](_page_37_Picture_6.jpeg)

### IUCN threat statuses synonyms sometimes have problems

×	Get data	How-to	Tools	Community	About		
<	Occur	rences	Ę	3			
	Тахо	on		TABLE	GALLERY	MAP	TA
Verbatim t	axon ID		~				
Scientific r	name		~				R
🔽 Bos tau	urus Linnaeus, 175	58				GR. P.	
Verbatim s	cientific name		~				
IUCN Glob	al Red List Catego	ry	^			-	
Z Extinct	(EX)		5,936	Bos p	primigenius sub	sp. taurus	A AN
Extinct	in the Wild (EW)		0	a second		ALL REAL PROPERTY	OW SAL
Critical	ly Endangered (CR	?)	0		1 the second	X	
Endang	gered (EN)		0		-	Super cash	Le Alle
🗌 Vulnera	able (VU)		0			L.u.	-
Near T	hreatened (NT)		0		the contract	-	-
Least C	Concern (LC)		0		-	in .	
Data D	eficient (DD)		0		all and a second		
Not Eva	aluated (NE)		412,525	Bos p	primigenius sub	osp. taurus	5-

METRICS

![](_page_38_Picture_5.jpeg)

![](_page_38_Picture_6.jpeg)

### **GBIF** uninterpreted free text fields

K	Get data	How-to	Tools	Соп	nmunity	About		
<	Occur	rences	Ţ	2				
Continent			~	•	TABLE	GALLERY	MAP	TA
Water body			~		+			
Island grou	р		~		-			
Island			~		<u>ک</u> ۲			
Clippert	ton				0			
Search								
CL	EAR							
Country or	area		~					
State provi	nce		~					
Locality			~					
Elevation			~		2 km			
Depth			~		© OpenStre	etMap contrib	utors, © Op	enMap

![](_page_39_Figure_2.jpeg)

![](_page_39_Picture_3.jpeg)

### Uninterpreted free text fields - State Province

K	Get data	How-to	Tools	Соп	nmunity	About		
<	Occur	rences						
Water body			~	-	TABLE	GALLERY	MAP	TA
Island group			~		+		85	
Island			~		-		S.	
Country or a	rea		~		7	64	22	8
State provin	ce		^			§ § Ø		
Search	I							Q
Locality			~				R	
Elevation			~					ė
Depth			~				•0	
Location			~					
Coordinate u	uncertainty in me	etres	~	Т	This map co	ontains occurre	ences flagg	ed by (
Georeferenc	ed by		~		OpenStre	etMap contribu	itors, © Op	enMap

- 🖓 🙀 🔍 📜 jwaller

SEARCH OCCURRENCES 2,959,730,060 WITH COORDINATES

XONOMY METRICS & DOWNLOAD

![](_page_40_Figure_5.jpeg)

![](_page_40_Picture_6.jpeg)

![](_page_40_Picture_7.jpeg)

![](_page_41_Figure_0.jpeg)

Access

![](_page_41_Picture_3.jpeg)

![](_page_42_Picture_2.jpeg)

#### What is a suspicious record?

- Example Geolocation outside of expected species range. Possible causes:
  - Label mistranscribed
  - Coordinate errors
  - Specimen misidentification
  - Etc.

https://github.com/gbif/portal-feedback/issues/5603

Synced a month ago

OCCURRENCE 12 JUNE 2004

#### Achillea oxyloba (DC.) Sch.Bip.

#### **Observed in Switzerland**

Plantae > Tracheophyta > Magnoliopsida > Asterales > Asteraceae > Achillea

DETAILS

Species: Achillea oxyloba (DC.) Sch.Bip.

Dataset: Schlern - (Bozen)

![](_page_43_Figure_15.jpeg)

![](_page_43_Picture_16.jpeg)

Taxon		2
Term	Interpreted	Original
Kingdom	Plantae	
Phylum	Tracheophyta	Chordata
Class	Magnoliopsida	Cephalaspic
Order	Myrtales	Anaspidifor
Family	Melastomataceae	
Genus	Chaetogastra	Pterolepis
Specific epithet	nitida	nitidus
Accepted name usage	Pterolepis nitidus	Pterolepis n
Generic name	Pterolepis	· · · · · · · · · · · · · · · · · · ·
Scientific name	Pterolepis nitida (Graham) Tr	iana
Rank	Species	
Taxonomic status	Synonym	

![](_page_44_Figure_3.jpeg)

- Verbatim vs interpreted view on GBIF
- Consider the basis of record
- Links to the source when available
- Download of the original archive published

![](_page_44_Picture_8.jpeg)

![](_page_45_Figure_1.jpeg)

- Observations
- Material sample
- Material citation
- Preserved specimen
- Fossil specimen
- Living specimen

- Verbatim vs interpreted view on GBIF
- Consider the basis of record
- Links to the source when available
- Download of the original archive published

![](_page_45_Picture_12.jpeg)

Synced 10 days ago

OCCURRENCE 2023 JANUARY Halomicroarcula sp. SYNS111 **Collected in China** Archaea > Halobacteriota > Halobacteria > Halobacteriales > Haloarculaceae DETAILS GBIF Taxon interpretation: Halomicroarcula Echigo et al., **Dataset: INSDC Sequences** 2013 **Publisher:** European Nucleotide Archive (EMBL-EBI) Location: Asia > China Reference: https://www.ebi.ac.uk/ena/brows-Basis of record: Material sample er/view/CP119795 **Issues:** Continent derived from country Taxon match higherrank

#### **Coordinates missing**

This record is published without coordinates, but it includes a textual description of its location.

Location: Asia > China Locality: Jiangsu

![](_page_46_Picture_6.jpeg)

https://www.gbif.org/occurrence/4155146316

![](_page_46_Picture_8.jpeg)

![](_page_46_Picture_9.jpeg)

- Verbatim vs interpreted view on GBIF
- Consider the **basis of record**
- Links to the **source** when available
- Download of the original archive published

![](_page_46_Picture_14.jpeg)

![](_page_47_Figure_1.jpeg)

science platform achieving real world outcomes through community. NatureMapr's mission is to empower anybody to report plant or animal information anywhere in Australia and ensure the information gets to the people that need to know about it. Values: Establish productive partnerships with local and state novernment Australia wide Allow hindiversity

Publication date: June 30, 2023 Metadata last modified: February 3, 2025 Hosted by: Atlas of Living Australia Licence: CC BY 4.0

- Verbatim vs interpreted view on GBIF
- Consider the **basis of record**
- Links to the **source** when available
- Download of the original archive published

![](_page_47_Picture_10.jpeg)

#### **Step 2: check the dataset metadata**

- Check metadata of the dataset:
  - Is it an ongoing project?
  - Was it a one-time initiative?
  - Is the source material available?
  - Is there other evidence available?

Get data How-to Tools Community About

#### - 🕂 🗙 Q 🔲 mgrosjean

#### Description

#### Description

Purpose

- Temporal scope
- Geographic scope
- Taxonomic scope
- Methodology
- Additional info
- Bibliography
- Contacts
- Data description
- **GBIF** registration
- Citation

A database with more than ten (10) years of information gathering and sampling in different localities of the country is presented. This database contains to date a total of 10.435 records of benthic macroalgae, with locations distributed along the Venezuelan coast. It is a compilation that has been made after the review, to date, of 208 bibliographic references, where a total of 657 species are reported, including 292 synonyms. The records of the Catalog Ficoflora Venezuela come from 500 georeferenced locations, distributed in 12 federal entities. Currently, three (3) special degree projects are also being developed within the framework of this project with the participation of students and teachers from the School of Computing and the School of Biology of the Faculty of Sciences of the Central University of Venezuela.

#### Purpose

Ficoflora Venezuela is a project carried out with the objective of updating the marine algae inventory at the national level, characterizing, morphoanatomically illustrating and taxonomically validating the available records, thus elaborating the first Digital Taxonomic Catalog of Marine Benthic Macroalgae of Venezuela.

#### Temporal scope

• January 1, 1847 - December 31, 2021

#### Geographic scope

![](_page_48_Picture_29.jpeg)

#### **Step 3: contact the dataset providers**

#### Record

Term	Interpreted	Original	Remar
Basis of record	Human observation	HumanObservation	
Collection code	Observations	Observations	
Dataset name	iNaturalist research-grade obsolvations	er- iNaturalist research-grade obser vations	-
Dynamic properties	{"evidenceOfPresence":"organ- ism"}	{"evidenceOfPresence":"organ- ism"}	
Institution code	iNaturalist	iNaturalist	
Occurrence			
Term	Interpreted	Original	Remar
Catalogue number	256870531	256870531	
Life stage	Adult	adult	
Occurrence ID	https://www.inatuva- tions/256870531	https://www.inatuva- tions/256870531	
Occurrence status	PRESENT		Inferred
Recorded by	Joseph Knight	Joseph Knight	
Sex	FEMALE	female	
Vitality	alive	alive	

![](_page_49_Figure_4.jpeg)

![](_page_49_Picture_5.jpeg)

#### **Step 3: contact the dataset providers**

#### National Museum of Natural History, Luxembourg

		3,122,034 OCCURRENCES	3,122,034 HOSTED OCCURRENCES	45 DATASETS	1,816 CITATIONS	
ABOUT	METRICS	A HOME PAGE				
Deserin	tions bits of					( <b>b</b>
Endorse	ad by: GBIE	/ www.mnnn.iu				
Installa	tions: Luxe	mbourg Metabarcoding	Data Toolkit •			
BioCAS	e Installatio	on Musée national d'histo	pire naturelle Luxembourg • IPT_M	NHNL		
Country	<b>v or area:</b> Lu	ixembourg				
Hosting	ı: 45 datase	ts ( 1 publisher∙ 1 coun	try )			
Down	load activity	/ report				
+ - Guernsey Jersey Jersey	Portsmouth Brigh	Con Calais A25 E40 A16 A25 Lile A16 A27 A A16 A27 A A16 A27 A A16 A27 A A16 A27 A A28 A13 A28 A14 A28 A15 A4 Paris	Brussels E429 Belgium E429 Belgium E429 Belgium E427 A34 Cologne A353 Bonn A3 Cologne A35 Bonn A3 Cologne A4 Cologne Cologne C	A45 A45 A45 A45 A45 A45 A45 A45 A45 A45	A4     Jena     A4       A9     A72       A9     A72       A9     A72       A9     A72       A9     A92       A71     A73       A9     A93       A73     A9       A33     A9       A6     A3       A9     Regensburg       A7     Ingolstadt       A9     A92       A9     A93	A 17 Prague C 2 Ref D4 C 3 Ref A 3 Linz
National	Museum o	f Natural History,	Patrick Michaely			
Luxembo	ourg		Head of delegation			

- Dataset contact information vs occurrence "recorded by"
- Contact institution
- GBIF feedback system

![](_page_50_Picture_7.jpeg)

#### Step 3: contact the dataset providers

![](_page_51_Picture_1.jpeg)

- Dataset contact information vs occurrence "recorded by"
- Contact institution
  - **GBIF feedback** system

![](_page_51_Picture_5.jpeg)

# Watch our previous sessions

datause@gbif.org

https://www.gbif.org/data-use-club

![](_page_52_Picture_3.jpeg)

![](_page_52_Picture_18.jpeg)