palaeoverse

Data acquisition

Bethany Allen & Harriet B. Drage

Objectives

1. Know some examples of different types of databases
2. Understand how to access raw data
3. Understand why it is important to keep raw data raw
4. Acquire and load example fossil dataset into R

Data acquisition: online databases

Many for different data types/origins/groups, e.g.:

- Geobiodiversity Database (stratigraphic approach; geobiodiversity.com)
- Neotoma (recent; neotomadb.org)
- iDigBio (mostly US museums; idigbio.org)
- Neptune/Triton (planktonic microfossils; nsb.mfn-berlin.de)
- BioDeepTime (time series; doi.org/10.1111/geb.13735)
- Phylacine (Quarternary mammals; megapast2future.github.io/)
- PARED (palaeo reefs; paleo-reefs.pal.uni-erlangen.de)

Geobiodiversity Database

Neotoma Paleoecology Database

iDigBio Integrated Digitized Biocollections

Neptune SB Berlin

PHYLACINE

paleoreefs

palaeoverse

Data acquisition: online databases

- IUCN: iucnredlist.org
- TreeBASE: treebase.org
- MorphoBank: morphobank.org
- Fossil Calibration Database: fossilcalibrations.org
- MorphoSource: morphosource.org
- Phenome10k: phenome10k.org
- Macrostrat: macrostrat.org
- EarthByte: earthbyte.org/category/resources/data-models/
- BRIDGE palaeoclimate models: bristol.ac.uk/geography/research/bridge
- CHELSA: chelsa-climate.org

Paleobiology Database (PBDB)

The Paleobiology Database
revealing the history of life

paleobiodb.org

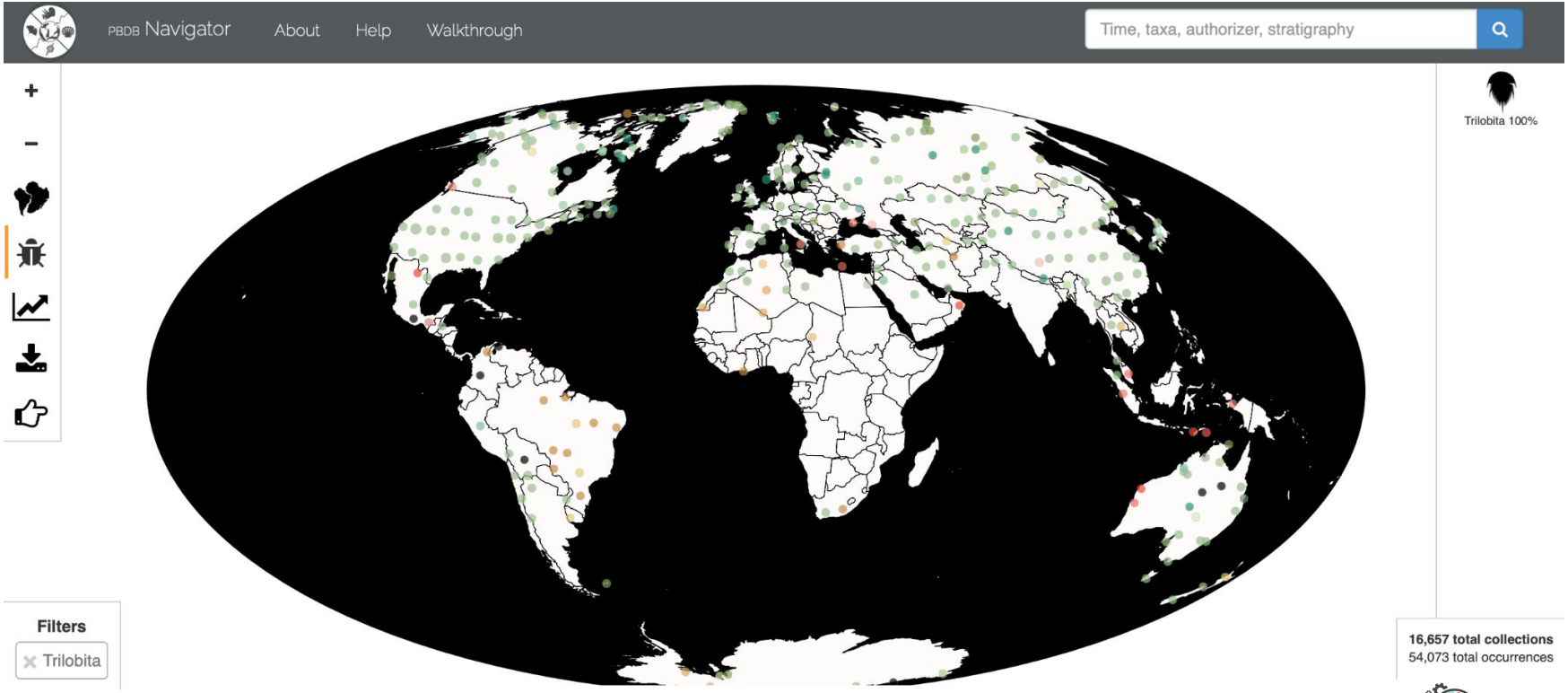
Over 20 years old, mostly funded by NSF

Coverage is global, good for macrofossils throughout geological time

Can be explored using the Navigator, but data can also be downloaded

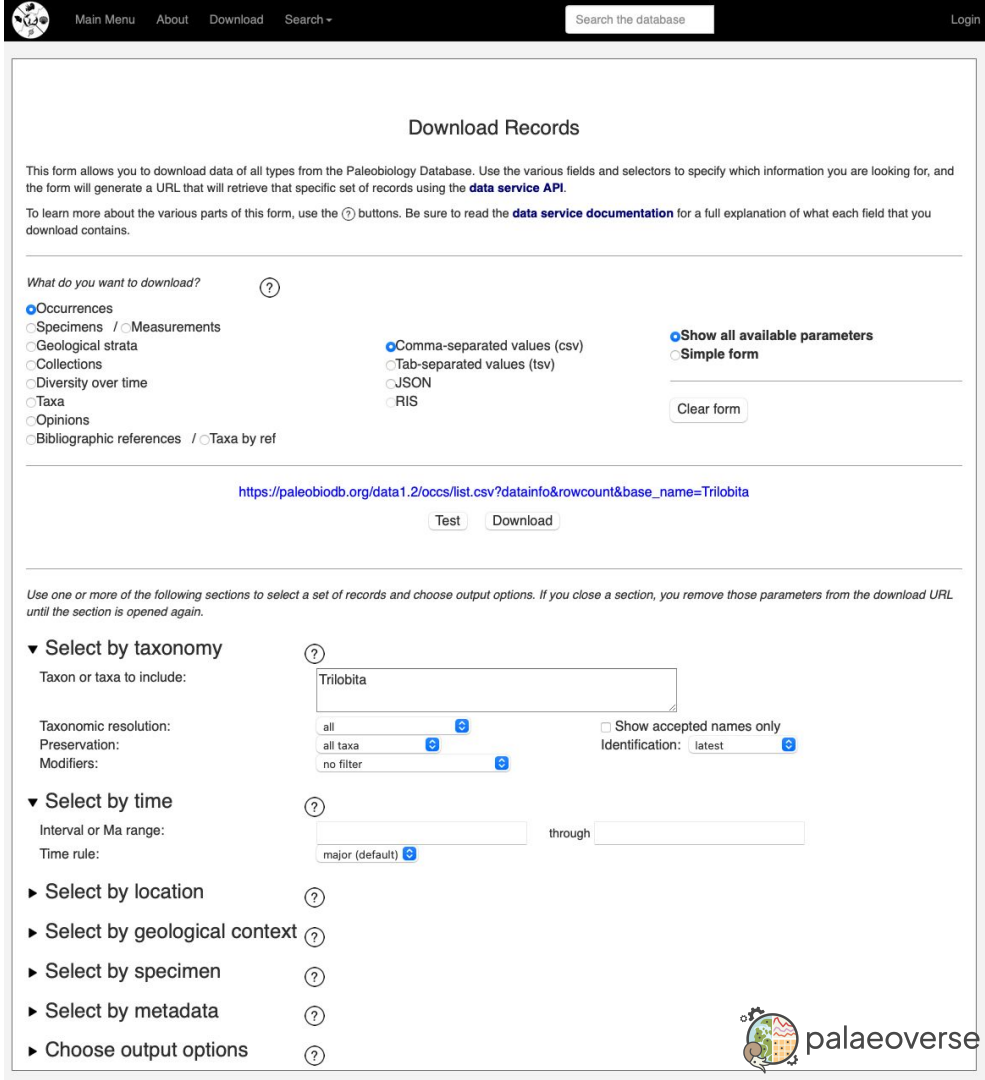
All fossils entered by palaeontologists from the published literature

PBDB user interface



palaeoverse

PBDB user interface



palaeoverse

PBDB user interface

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Data Provide	The Paleobiology Database															
2	Data Source	The Paleobiology Database															
3	Data License	Creative Commons CC0															
4	License URL	https://creativecommons.org/publicdomain/zero/1.0/															
5	Documentat	http://paleobiodb.org/data1.2/occs/list_doc.html															
6	Data URL	http://paleobiodb.org/data1.2/occs/list.csv?datainfo&rowcount&base_name=Trilobita															
7	Access Time	Thu 2025-07-03 13:54:09 GMT															
8	Title	PBDB Data Service															
9	Parameters:																
10		base_name	Trilobita														
11		timerule	major														
12		taxon_status	all														
13	Elapsed Time	3.84															
14	Records Fou	54073															
15	Records Retu	54073															
16	Records:																
17	occurrence_	record_type	reid_no	flags	collection_n	identified_na	identified_ra	identified_no	difference	accepted_na	accepted_ra	accepted_no	early_interva	late_interval	max_ma	min_ma	reference_no
18	1	occ			1	Australosutu	species	349412		Australosutu	species	349412	Ivorian		353.7	346.7	1
19	2	occ			1	Carbonocory	species	349526	recombined a	Phillibole pla	species	349526	Ivorian		353.7	346.7	1
20	3	occ			1	Thigriffides rc	species	349420		Thigriffides rc	species	349418	Ivorian		353.7	346.7	1

palaeoverse

Global Biodiversity Information Facility (GBIF)

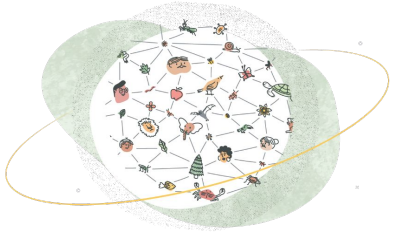
gbif.org

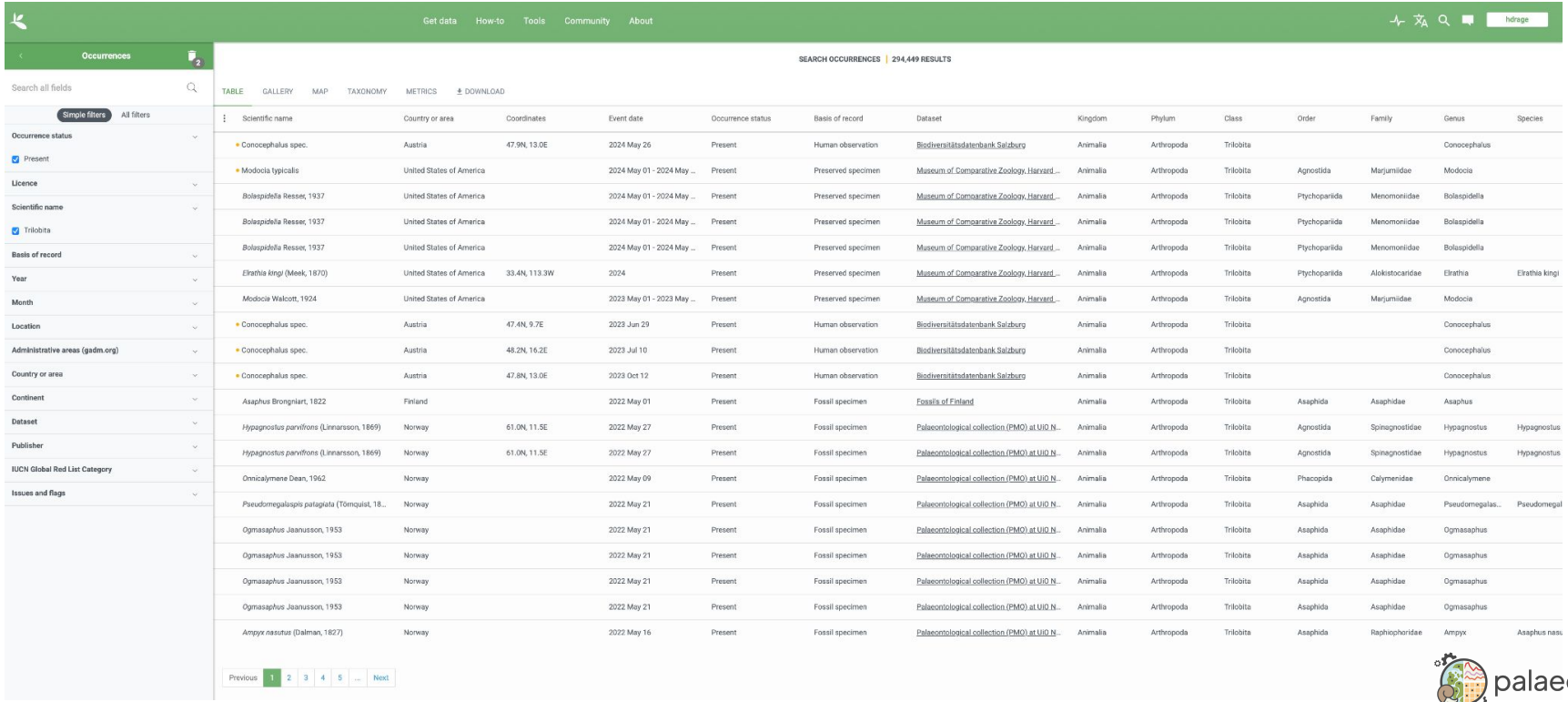
Compiles modern and fossil occurrence data from a wide variety of sources, including direct reporting from museum collections

(includes PBDB data)

Also has extensive taxonomic records

GBIF user interface





Occurrence status

☑ Present

Licence

Scientific name

☑ Trilobita

Basis of record

Year

Month

Location

Administrative areas (gadm.org)

Country or area

Continent

Dataset

Publisher

IUCN Global Red List Category

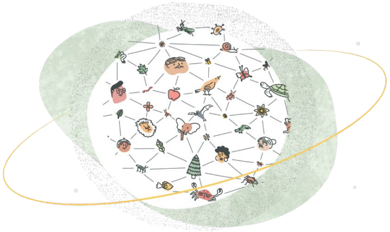
Issues and flags

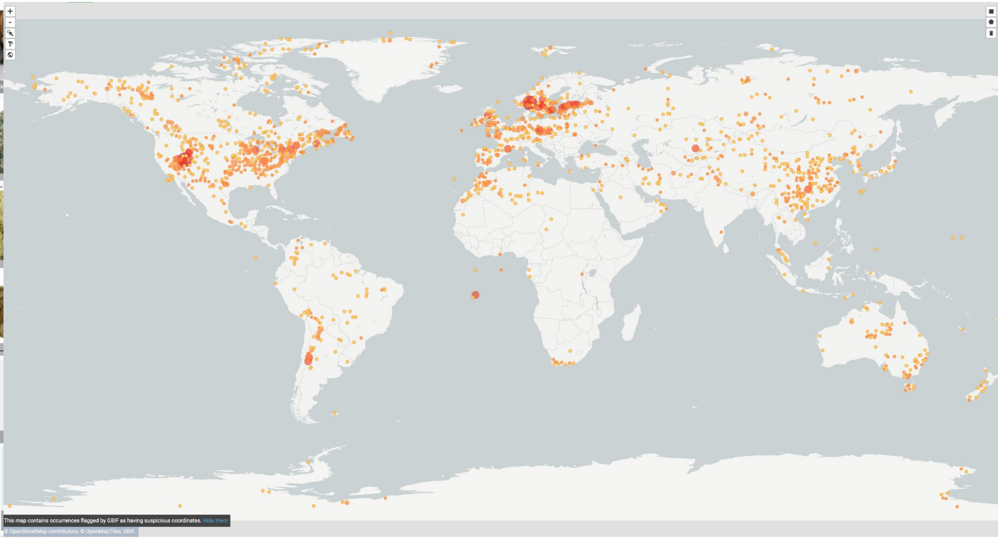
SEARCH OCCURRENCES | 294,449 RESULTS

Scientific name	Country or area	Coordinates	Event date	Occurrence status	Basis of record	Dataset	Kingdom	Phylum	Class	Order	Family	Genus	Species
• Conocephalus spec.	Austria	47.9N, 13.0E	2024 May 26	Present	Human observation	Biodiversitätsdatenbank Salzburg	Animalia	Arthropoda	Trilobita			Conocephalus	
• Modocia typicalis	United States of America		2024 May 01 - 2024 May ...	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Agnostida	Marjumiidae	Modocia	
Bolaspidella Resser, 1937	United States of America		2024 May 01 - 2024 May ...	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Ptychopariida	Menomoniidae	Bolaspidella	
Bolaspidella Resser, 1937	United States of America		2024 May 01 - 2024 May ...	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Ptychopariida	Menomoniidae	Bolaspidella	
Bolaspidella Resser, 1937	United States of America		2024 May 01 - 2024 May ...	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Ptychopariida	Menomoniidae	Bolaspidella	
Elrathia kingi (Meek, 1870)	United States of America	33.4N, 113.3W	2024	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Ptychopariida	Alokistocaridae	Elrathia	Elrathia kingi
Modocia Walcott, 1924	United States of America		2023 May 01 - 2023 May ...	Present	Preserved specimen	Museum of Comparative Zoology, Harvard ...	Animalia	Arthropoda	Trilobita	Agnostida	Marjumiidae	Modocia	
• Conocephalus spec.	Austria	47.4N, 9.7E	2023 Jun 29	Present	Human observation	Biodiversitätsdatenbank Salzburg	Animalia	Arthropoda	Trilobita			Conocephalus	
• Conocephalus spec.	Austria	48.2N, 16.2E	2023 Jul 10	Present	Human observation	Biodiversitätsdatenbank Salzburg	Animalia	Arthropoda	Trilobita			Conocephalus	
• Conocephalus spec.	Austria	47.8N, 13.0E	2023 Oct 12	Present	Human observation	Biodiversitätsdatenbank Salzburg	Animalia	Arthropoda	Trilobita			Conocephalus	
Asaphus Brongniart, 1822	Finland		2022 May 01	Present	Fossil specimen	Fossils of Finland	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Asaphus	
Hypagnostus parvifrons (Linnarsson, 1869)	Norway	61.0N, 11.5E	2022 May 27	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Agnostida	Spinagnostidae	Hypagnostus	Hypagnostus
Hypagnostus parvifrons (Linnarsson, 1869)	Norway	61.0N, 11.5E	2022 May 27	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Agnostida	Spinagnostidae	Hypagnostus	Hypagnostus
Onnicalymene Dean, 1962	Norway		2022 May 09	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Phacopida	Calymenidae	Onnicalymene	
Pseudomegalaspis patagiata (Törnquist, 18...	Norway		2022 May 21	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Pseudomegalas...	Pseudomegal
Ogmasaphus Jaanusson, 1953	Norway		2022 May 21	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Ogmasaphus	
Ogmasaphus Jaanusson, 1953	Norway		2022 May 21	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Ogmasaphus	
Ogmasaphus Jaanusson, 1953	Norway		2022 May 21	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Ogmasaphus	
Ogmasaphus Jaanusson, 1953	Norway		2022 May 21	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Asaphidae	Ogmasaphus	
Ampyx nasutus (Dalman, 1827)	Norway		2022 May 16	Present	Fossil specimen	Palaeontological collection (PMO) at UiO N...	Animalia	Arthropoda	Trilobita	Asaphida	Raphiophoridae	Ampyx	Asaphus nasu

palaeoverse

GBIF user interface



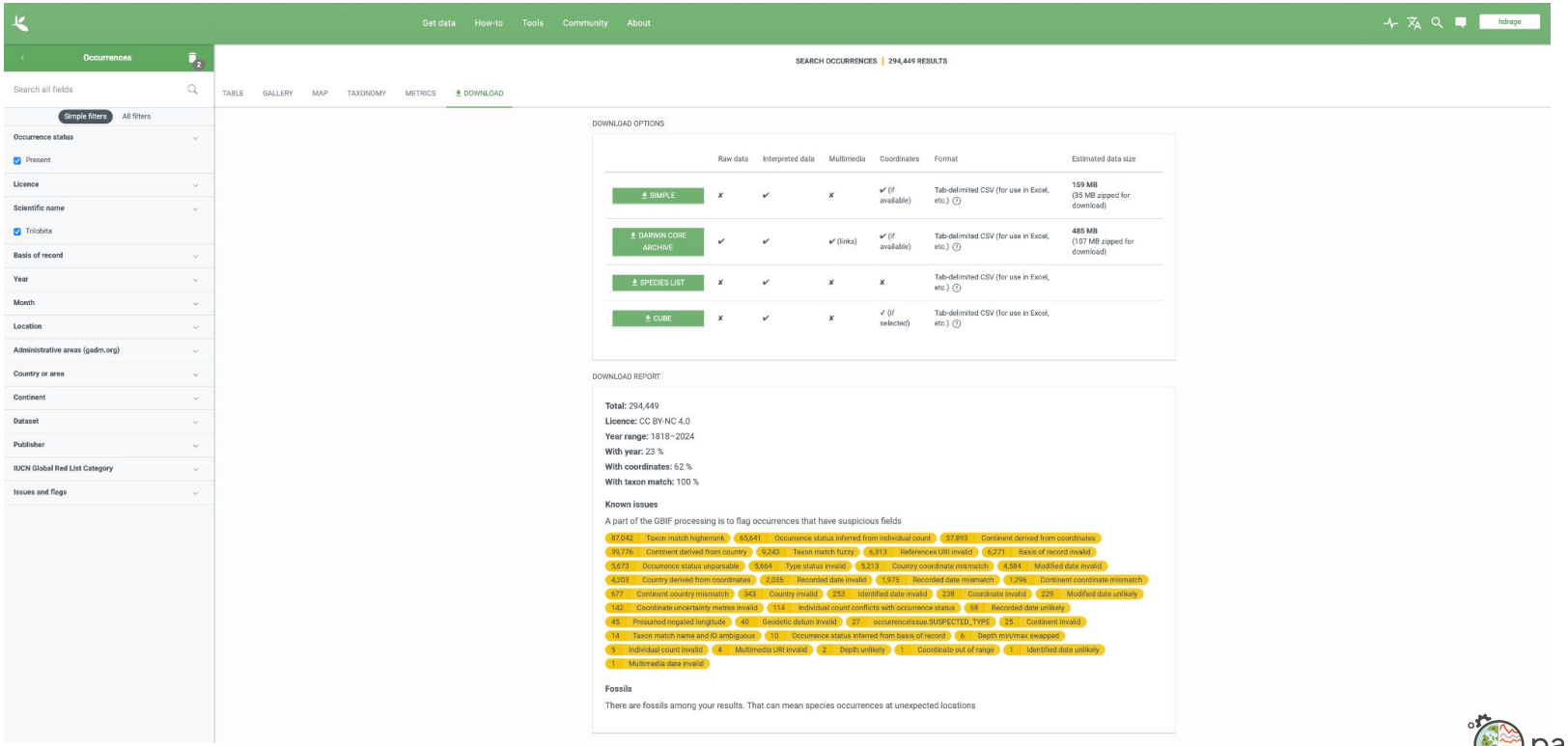


This map contains occurrences flagged by GBIF as having suspicious coordinates. Hide them

© OpenStreetMap contributors, © OpenMapTiles, GBIF.

palaeoverse

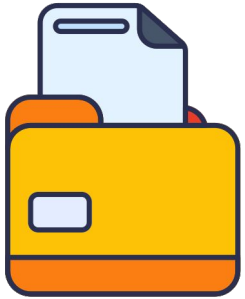
GBIF user interface



palaeoverse

What is raw data and why keep it raw?

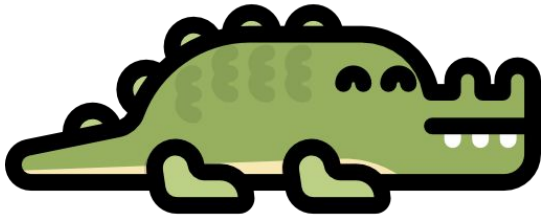
- **The data you originally downloaded - no changes!**
- Why?:
  - identification of later errors
  - reproducibility
  - online databases are not static
- How?
  - store files locally and clearly
  - check file encoding
  - read-only - copy file to make changes
  - clean using R/other language



Today’s research question

How does range of Crocodylia change across the Cenozoic with environmental conditions?

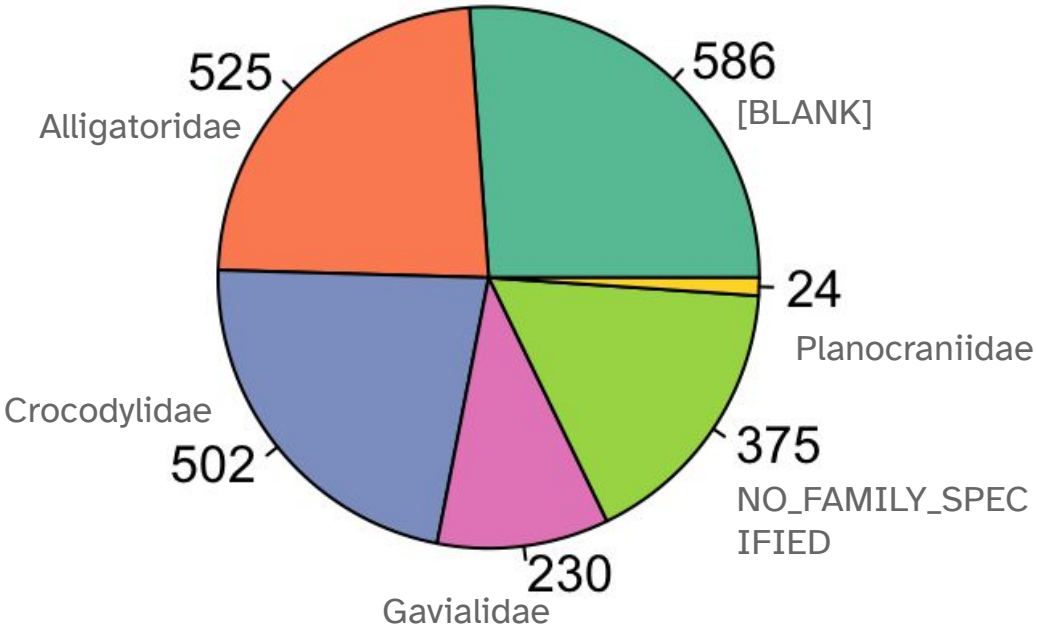
Crocs are a good group to look at for palaeo/ecology - fossil record, modern data, responsive to temperature, global record

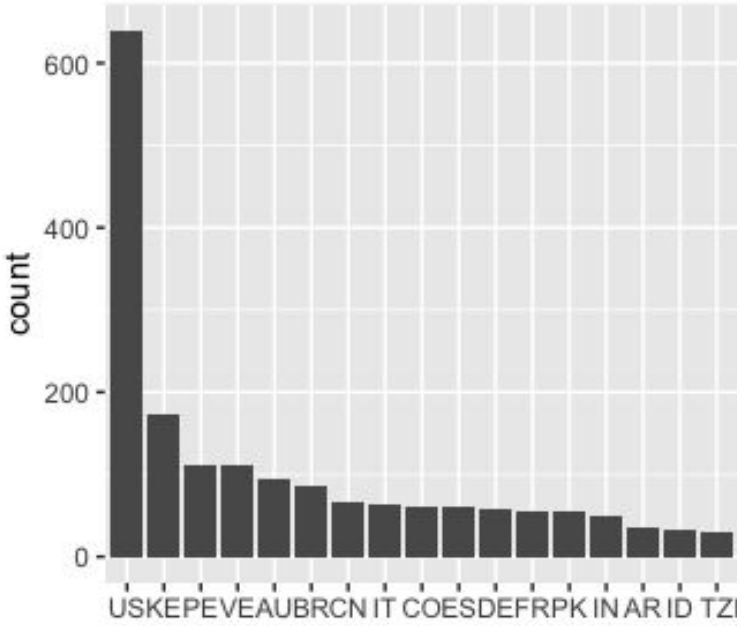


palaeoverse

Let’s load our data!

Our palaeo dataset





palaeoverse

On to data processing!