Topic: Working session for taxonomic data, Part I
Date: 2022 February 8
Time: 12 PM Eastern

In Attendance

Erica Krimmel, iDigBio	Talia Karim, CUMNH	Jessica Cundiff, MCZ
Jessica Utrup, YPM	Roger Burkhalter, OMNH	Holly Little, NMNH
Suzanne Mills, CarMNH	Vicky Wang, PRI	Jacob Van Veldhuizen, CUMNH
Amanda Millhouse, NMNH	Christina Byrd, MCZ	Jeremy Wyman, VA DOE
Lindsay Walker	Gil Nelson, iDigBio	Kari Lintulaakso, NHM Helsinki

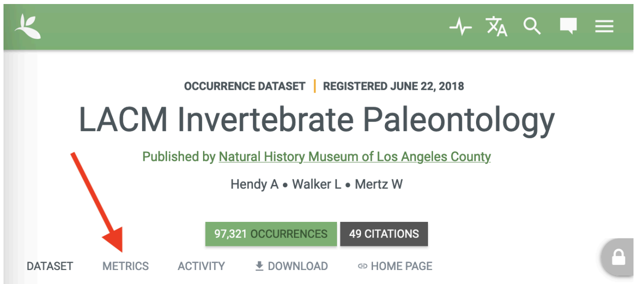
Resources

- Paleo Data Working Group Google Folder
- Paleo Data Working Group wiki page with schedule of upcoming Happy Hours

Homework

Identify issues with paleo taxonomy on GBIF by exploring data in one of the following ways:

A. ***Look at the data quality flags assigned to your collection by GBIF during data ingestion.***
    1. Find the dataset page on GBIF for your collection, or, if you are not publishing data to GBIF yet, look up another collection you are familiar with. E.g., here is the dataset page for LACMIP: https://www.gbif.org/dataset/f0a7ca6e-8da6-4629-97bd-0368705a4d6b
    2. From the dataset page, click on the Metrics tab.



OCCURRENCE DATASET | REGISTERED JUNE 22, 2018

LACM Invertebrate Paleontology

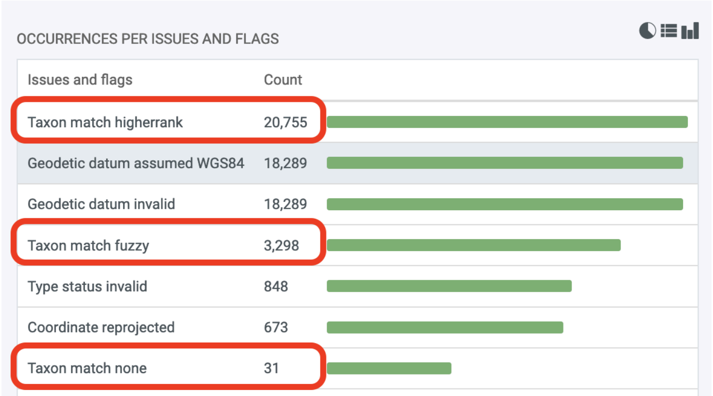
Published by Natural History Museum of Los Angeles County

Hendy A • Walker L • Mertz W

97,321 OCCURRENCES

49 CITATIONS

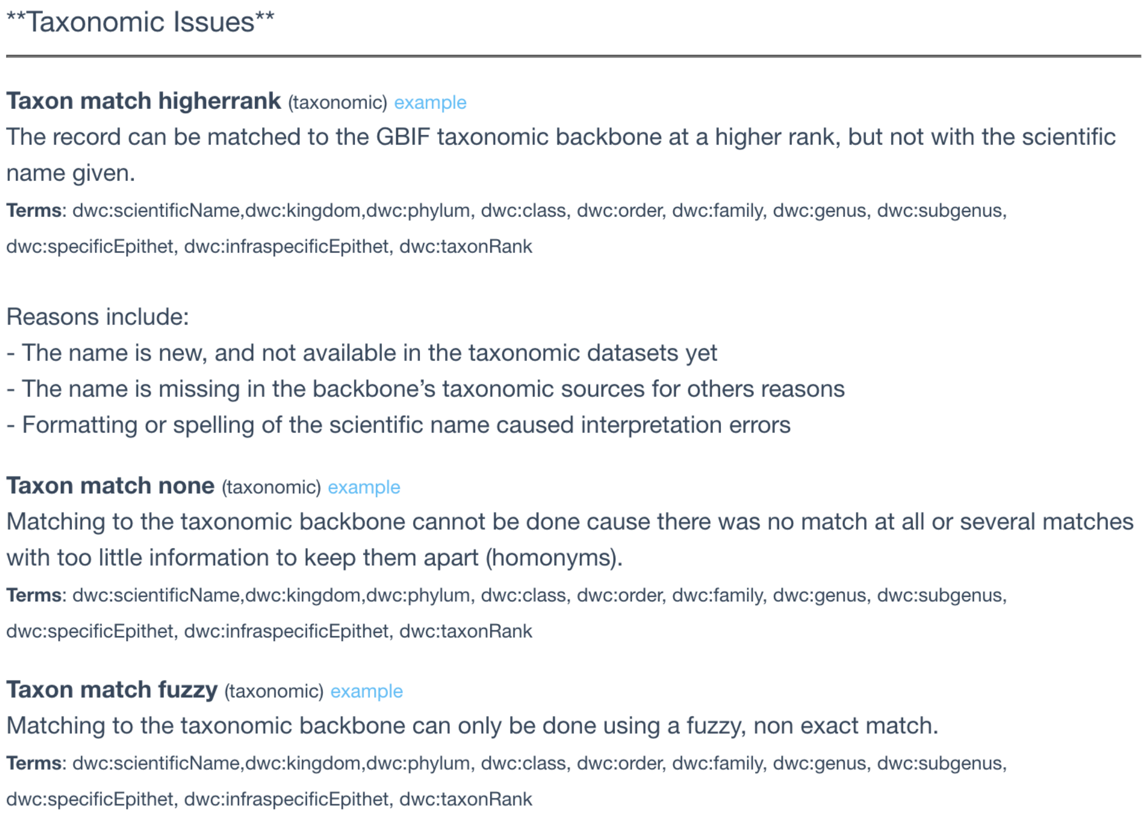
    3. Scroll down to the bottom left of the Metrics tab where there is a box labeled, Occurrences per issues and flags. Look for flags that relate to taxonomy, like “Taxon match fuzzy.” To see the specimen records that have been assigned this flag, click on the flag of interest.



OCCURRENCES PER ISSUES AND FLAGS

Issues and flags	Count
Taxon match higherrank	20,755
Geodetic datum assumed WGS84	18,289
Geodetic datum invalid	18,289
Taxon match fuzzy	3,298
Type status invalid	848
Coordinate reprojected	673
Taxon match none	31

Taxonomic flags, as described at https://data-blog.gbif.org/post/issues-and-flags/:



Taxonomic Issues

Taxon match higherrank (taxonomic) example

The record can be matched to the GBIF taxonomic backbone at a higher rank, but not with the scientific name given.

Terms: dwc:scientificName,dwc:kingdom,dwc:phylum, dwc:class, dwc:order, dwc:family, dwc:genus, dwc:subgenus, dwc:specificEpithet, dwc:infraspecificEpithet, dwc:taxonRank

Reasons include:
- The name is new, and not available in the taxonomic datasets yet
- The name is missing in the backbone’s taxonomic sources for others reasons
- Formatting or spelling of the scientific name caused interpretation errors

Taxon match none (taxonomic) example

Matching to the taxonomic backbone cannot be done cause there was no match at all or several matches with too little information to keep them apart (homonyms).

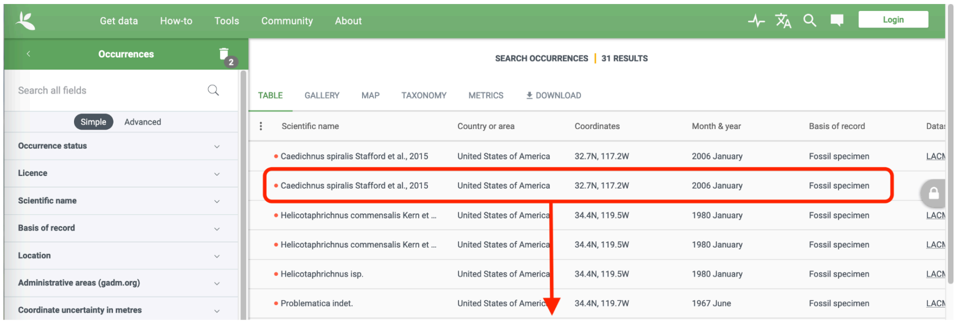
Terms: dwc:scientificName,dwc:kingdom,dwc:phylum, dwc:class, dwc:order, dwc:family, dwc:genus, dwc:subgenus, dwc:specificEpithet, dwc:infraspecificEpithet, dwc:taxonRank

Taxon match fuzzy (taxonomic) example

Matching to the taxonomic backbone can only be done using a fuzzy, non exact match.

Terms: dwc:scientificName,dwc:kingdom,dwc:phylum, dwc:class, dwc:order, dwc:family, dwc:genus, dwc:subgenus, dwc:specificEpithet, dwc:infraspecificEpithet, dwc:taxonRank

4. When you click on a flag from the box above, it will bring you to the GBIF occurrence search interface and display a list of records assigned the flag of interest. Choose one of these records and click on it to get to the occurrence (specimen) record page, where you will be able to see data that will help you understand why the flag was assigned.



SEARCH OCCURRENCES | 31 RESULTS

Scientific name	Country or area	Coordinates	Month & year	Basis of record
Caedichnus spiralis Stafford et al., 2015	United States of America	32.7N, 117.2W	2006 January	Fossil specimen
Caedichnus spiralis Stafford et al., 2015	United States of America	32.7N, 117.2W	2006 January	Fossil specimen
Helicotaphrichnus commensalis Kern et ...	United States of America	34.4N, 119.5W	1980 January	Fossil specimen
Helicotaphrichnus commensalis Kern et ...	United States of America	34.4N, 119.5W	1980 January	Fossil specimen
Helicotaphrichnus isp.	United States of America	34.4N, 119.5W	1980 January	Fossil specimen
Problematica indet.	United States of America	34.4N, 119.7W	1967 June	Fossil specimen

Synced a year ago
Modified 4 June 2020

OCCURRENCE | 1 JANUARY 2006

Caedichnus spiralis Stafford et al., 2015

Collected in United States of America

GBIF Taxon interpretation: none
Kingdom: incertae sedis
Location: North America › United States of America
Basis of record: Fossil specimen

Dataset: LACM Invertebrate Paleontology
Publisher: Natural History Museum of Los Angeles County
Issues: Taxon match none

5. On the occurrence (specimen) record page, scroll down to the section labeled, Taxon. This is where GBIF displays–from left to right, for each field–the Darwin Core **Term** describing what data should be in this field, the value for this field that GBIF has **Interpreted** as being "correct," the **Original** value that you had in this field, and Remarks about GBIF's data ingestion process. In the **Remarks** you will see data flags, e.g. "Taxon match none." Look carefully at the difference between GBIF's Interpreted (outlined in red below) and your Original data. Is GBIF's interpretation correct? If not, can you tell why they misinterpreted your data?

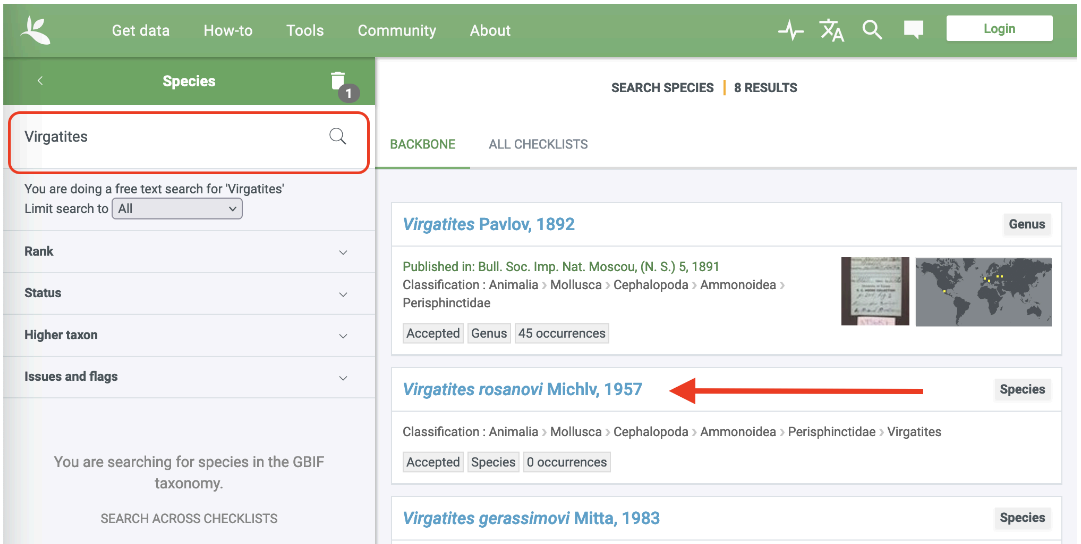
Taxon

Term	Interpreted	Original	Remarks
Kingdom	incertae sedis	Ichnotaxa	Taxon match none
Phylum		Incertae sedis	Taxon match none
Class		Incertae sedis	Taxon match none
Order		Incertae sedis	Taxon match none
Family		Incertae sedis	Taxon match none
Genus		Caedichnus	Taxon match none
Specific epithet		spiralis	Taxon match none
Higher classification	Ichnotaxa | Caedichnus	Ichnotaxa | Caedichnus	
Scientific name	incertae sedis	Caedichnus spiralis Stafford et al., 2015	Taxon match none
Scientific name authorship		Stafford et al., 2015	Taxon match none
Rank	Kingdom	Species	Altered

6. Make a note of the instances where GBIF's interpretation is *not* correct. If possible, save links to occurrence (specimen) records that can serve as examples. You can add your notes/links to this spreadsheet.

B. *Look up a taxon that you suspect GBIF may have trouble interpreting.*

1. Go to https://www.gbif.org/species/search and type in the name of the taxa you want to look up in the left hand bar. View results from the GBIF Taxonomic Backbone in the right hand display, and click on any taxa to see more details.



SEARCH SPECIES | 8 RESULTS

Virgatites

You are doing a free text search for 'Virgatites'
Limit search to All

You are searching for species in the GBIF taxonomy.

SEARCH ACROSS CHECKLISTS

Virgatites Pavlov, 1892
Published in: Bull. Soc. Imp. Nat. Moscou, (N. S.) 5, 1891
Classification : Animalia › Mollusca › Cephalopoda › Ammonoidea › Perisphinctidae
Accepted | Genus | 45 occurrences

Virgatites rosanovi Michlv, 1957
Classification : Animalia › Mollusca › Cephalopoda › Ammonoidea › Perisphinctidae › Virgatites
Accepted | Species | 0 occurrences

Virgatites gerassimovi Mitta, 1983

2. Is the taxon you are looking up represented the way you think it should be, i.e. with the same higher taxonomy and status (accepted/synonym) you would use? Make a note of the instances where GBIF's interpretation is *not* correct. If possible, save links to species records that can serve as examples. You can add your notes/links to this spreadsheet.

Agenda

- Review issues with paleo taxonomy in GBIF (homework)
  - See spreadsheet with specific examples and slide deck with synthesized examples.
- Best practices for recording and sharing taxonomic data
- Next steps for taxonomy stuff
  - Talk to GBIF.
  - Improve our data in house.
  - Propose a GBIF Hosted Portal for fossil specimens? Could be similar to the Legume Data Portal.

Action Items

- Add hard copy sources to taxonomic authorities Google doc
  - *Treatise*
- Review best practices for recording paleo taxonomy, e.g. locally in institutional database
  - Best taxonomic authorities for various groups. See and add to the list we have started here.
  - Need to align ourselves on how we record taxa with unknown higher classification.
  - Need to align ourselves on how we record unranked clades.

    - Need to align ourselves on this use of identification qualifiers, e. g. “Indet.” or “cf.”
- Review next practices for sharing paleo taxonomy, e.g. via GBIF
    - Best practices for using Darwin Core terms related to taxonomy and identifications. See our group’s working documentation here.
    - Need to align ourselves on how we share data about taxa with unknown higher classification. See GitHub issue here for examples related to populating dwc:higherclassificiation.
    - Need to determine what the highest priority interstitial rank terms we want. E.g. Suborder, Subclass. Note that subfamily was added as a term to Darwin Core in July 2021.

Notes

- Suzanne’s example: https://www.gbif.org/occurrence/1272006679
- New term: dwc:genus vs. dwc:genericName
    - https://github.com/tdwg/dwc/issues/29
    - https://dwc.tdwg.org/terms/#dwc:genericName
    - = Dumping ground for whatever the lowest rank provided is? (not necessarily genus?)
- Many major dinosaur orders, early mammals, excluded from many VP datasets (e.g. Saurischia, Avetheropoda)
- Backbone issues
    - How to indicate when/if PBDB “trumps” the taxonomy being used for the GBIF backbone from other resources, e.g. CoL?
    - How often is the Catalog of Life backbone refreshed?
    - PBDB taxonomy last refreshed in GBIF Nov 2021: https://www.gbif.org/dataset/c33ce2f2-c3cc-43a5-a380-fe4526d63650
    - Example of species added to PBDB and ZooBank in 2019, was apparently pulled from ZB: https://www.gbif.org/species/11119311 (took awhile)

Chat Copy

09:07:44 From Vicky Wang (V, V's / no pronouns) to Everyone:
Eurypteris remipes
09:08:02 From Vicky Wang (V, V's / no pronouns) to Everyone:
(sorry!)
09:08:24 From Jessica Utrup to Everyone:
No worries! I've just catalogued way too many!
09:10:15 From Talia S. Karim (She/Her) to Everyone:
https://data-blog.gbif.org/post/issues-and-flags/
09:10:33 From Lindsay Walker to Everyone:
This is really helpful… I did not know about this post!
09:12:20 From Christina J Byrd to Everyone:
Can someone please share the link to the meeting notes?

09:12:28 From Jessica Utrup to Everyone:
https://docs.google.com/document/d/1PNHvA4Fy7_-PbD4-rEM1psAXTJjNUp6a23bzo7MXSSo/edit?usp=sharing
09:12:39 From Suzanne Mills to Everyone:
https://www.gbif.org/occurrence/1272006679
09:12:45 From Christina J Byrd to Everyone:
Thank you
09:13:43 From Amanda Millhouse to Everyone:
GBIF is struggling with a lot of VP orders….
09:15:16 From Lindsay Walker to Everyone:
Holly, if you have the link to the issue discussion handy, can you share it?
09:15:36 From Holly Little to Everyone:
I can find it. But looking at the definition it does sound like it should still be Genus related
09:16:24 From Holly Little to Everyone:
Here’s the dwc term details https://dwc.tdwg.org/terms/#dwc:genericName
09:17:42 From Holly Little to Everyone:
here’s the GitHub thread that got us the term: https://github.com/tdwg/dwc/issues/29
09:20:56 From Holly Little to Everyone:
NMNH has examples like that too
09:21:00 From Vicky Wang (V, V's / no pronouns) to Everyone:
PRI sees this a lot with subspecies raised to species in GBIF.
09:22:28 From Jessica Utrup to Everyone:
If you change Auchenorrhyncha, please let me know. We use it quite a bit too as a suborder. Also, Ammonites virgatus is the genotype of Virgatites.
09:23:25 From Christina J Byrd to Everyone:
Yay dinosaur taxonomy!
09:33:03 From Suzanne Mills to Everyone:
am I interpreting correctly that we really don't know what the GBIF taxonomy backbone is?
09:34:07 From Erica Krimmel to Everyone:
I think functionally, not really. We can look up any name and see where the GBIF Backbone source for that name is, but as a whole it's difficult to understand what their logic is for when to use what taxonomic authority.
09:34:26 From Lindsay Walker to Everyone:
https://www.gbif.org/species/11119311
09:34:29 From Holly Little to Everyone:
General overview for the backbone https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c
09:34:30 From Suzanne Mills to Everyone:
well put Erica, that's what I'm understanding, too.
09:35:38 From Suzanne Mills to Everyone:
Thanks, Holly. "Additional scientific names only found in other authoritative nomenclatural and taxonomic datasets" added to Catalogue of Life could use some more explanation.
09:35:55 From Suzanne Mills to Everyone:

*added to Catalogue of Life backbone
09:37:01 From Holly Little to Everyone:
https://www.gbif.org/dataset/c33ce2f2-c3cc-43a5-a380-fe4526d63650
09:40:38 From Jessica Utrup to Everyone:
Am I just really old-fashioned? We use the Treatise on Invertebrate Paleontology to arrange the fossils physically in our collection.
09:41:42 From Suzanne Mills to Everyone:
We reference the Treatise a lot, too
09:41:54 From Lindsay Walker to Everyone:
@Jessica at LACMIP we used the Treatise selectively, e.g. just for ammonoids, but we do reference it as well
09:44:07 From Jessica Utrup to Everyone:
@Lindsay and @Suzanne, our systematic collection is arranged following the Treatise. We don't move things when a genus gets moved to another family unless a researcher comes in and corrects it - we just don't have the time/staff and it gives us a way to find the specimens. Often when we're adding taxonomy to the database, the species is new and we use the paper.
09:45:27 From Jessica D. Cundiff to Everyone:
We also use the treatise for physically arranging the collection and use Yale's taxonomic table in our database.
09:46:05 From Suzanne Mills to Everyone:
I appreciate hearing everyone's uses of the Treatise.
09:52:01 From Lindsay Walker to Everyone:
Do we care about best practices for separators? (e.g. pipes, commas, etc) … too in the weeds?
09:52:17 From Lindsay Walker to Everyone:
Animalia | Annelida | Polychaeta | Sedentaria | Sabellida
09:54:18 From Holly Little to Everyone:
I added a “Questions for GBIF” section in the sharing_taxonomy_working-document for brainstorming
09:56:18 From Lindsay Walker to Everyone:
Other examples of new paleo-focused portals
09:56:30 From Lindsay Walker to Everyone:
https://serv.biokic.asu.edu/paleo/portal/
09:56:37 From Erica Krimmel to Everyone:
👍
09:58:07 From Lindsay Walker to Everyone:
+1 more taxonomy
10:00:58 From Holly Little to Everyone:
I added a section to start that list too in the sharing_taxonomy document