A Review of Biotic Interactions and Taxon Names Found in globalbioticinteractions/yale-peabody hash://md5/41c7a250499cb3081abba26b82463ad9

by Nomer, Elton and Preston, three naive review bots
review@globalbioticinteractions.org
https://globalbioticinteractions.org/contribute
https://github.com/globalbioticinteractions/yale-peabody/issues

2025-04-30

Abstract

Life on Earth is sustained by complex interactions between organisms and their environment. These biotic interactions can be captured in datasets and published digitally. We present a review and archiving process for such an openly accessible digital interactions dataset of known origin and discuss its outcome. The dataset under review, named globalbioticinteractions/yale-peabody, has fingerprint hash://md5/41c7a250499cb3081abba26b82463ad9, is 466MiB in size and contains 283,454 interaction with 9 unique types of associations (e.g., adjacentTo) between 8,205 primary taxon (e.g., undet. Foraminiferida) and 140,237 associated taxon (e.g., rocks). This report includes detailed summaries of interaction data, a taxonomic review from multiple catalogs, and an archived version of the dataset from which the reviews are derived.

Contents

1

Discussion   **24**

Acknowledgements   **25**

Author contributions   **25**

References   **25**

Introduction

Data Review and Archive

Data review and archiving can be a time-consuming process, especially when done manually. This review report aims to help facilitate both activities. It automates the archiving of datasets, including Darwin Core archives, and is a citable backup of a version of the dataset. Additionally, an automatic review of species interaction claims made in the dataset is generated and registered with Global Biotic Interactions (J. H. Poelen, Simons, and Mungall 2014).

This review includes summary statistics about, and observations about, the dataset under review:

Botany Division, Yale Peabody Museum - Version 396.3287 http://ipt.peabody.yale.edu/archive.do?r=ypm-bot 2025-04-26T07:06:49.040Z Entomology Division, Yale Peabody Museum - Version 395.3193 http://ipt.peabody.yale.edu/archive.do?r=ypm-ent 2025-04-26T07:06:49.040Z Invertebrate Paleontology Division, Yale Peabody Museum - Version 395.3225 http://ipt.peabody.yale.edu/archive.do?r=ypm-ip 2025-04-26T07:06:49.040Z Invertebrate Zoology Division, Yale Peabody Museum - Version 395.3292 http://ipt.peabody.yale.edu/archive.do?r=ypm-iz 2025-04-26T07:06:49.040Z Paleobotany Division, Yale Peabody Museum - Version 395.3269 http://ipt.peabody.yale.edu/archive.do?r=ypm-pb 2025-04-26T07:06:49.040Z Vertebrate Paleontology Division, Yale Peabody Museum - Version 395.3290 http://ipt.peabody.yale.edu/archive.do?r=ypm-vp 2025-04-26T07:06:49.040Z Vertebrate Zoology Division - Herpetology, Yale Peabody Museum - Version 395.3254 http://ipt.peabody.yale.edu/archive.do?r=ypm-vz-her 2025-04-26T07:06:49.040Z Vertebrate Zoology Division - Ichthyology, Yale Peabody Museum - Version 395.3208 http://ipt.peabody.yale.edu/archive.do?r=ypm-vz-ich 2025-04-26T07:06:49.040Z Vertebrate Zoology Division - Mammalogy, Yale Peabody Museum - Version 397.3265 http://ipt.peabody.yale.edu/archive.do?r=ypm-vz-mam 2025-04-26T07:06:49.040Z Vertebrate Zoology Division - Ornithology, Yale

For additional metadata related to this dataset, please visit https://github.com /globalbioticinteractions/yale-peabody and inspect associated metadata files including, but not limited to, *README.md*, *eml.xml*, and/or *globi.json*.

Methods

The review is performed through programmatic scripts that leverage tools like Preston (Elliott et al. 2025), Elton (Kuhn, Poelen, and Leinweber 2025), Nomer (Salim and Poelen 2025), globinizer (J. Poelen, Seltmann, and Mietchen 2024) combined with third-party tools like grep, mlr, tail and head.

Table 1: Tools used in this review process

tool name	version
preston	0.10.1
elton	0.15.9
nomer	0.5.13
globinizer	0.4.0
mlr	6.0.0
jq	1.6
yq	4.25.3
pandoc	3.1.6.1

The review process can be described in the form of the script below [1].

```
# get versioned copy of the dataset (size approx.  466MiB) under review
elton pull globalbioticinteractions/yale-peabody

# generate review notes
elton review globalbioticinteractions/yale-peabody\
 > review.tsv

# export indexed interaction records
elton interactions globalbioticinteractions/yale-peabody\
 > interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names globalbioticinteractions/yale-peabody\
```

---

[1] Note that you have to first get the data (e.g., via elton pull globalbioticinteractions/yale-peabody) before being able to generate reviews (e.g., elton review globalbioticinteractions/yale-peabody), extract interaction claims (e.g., elton interactions globalbioticinteractions/yale-peabody), or list taxonomic names (e.g., elton names globalbioticinteractions/yale-peabody)

```
  | nomer append col\
> name-alignment.tsv
```
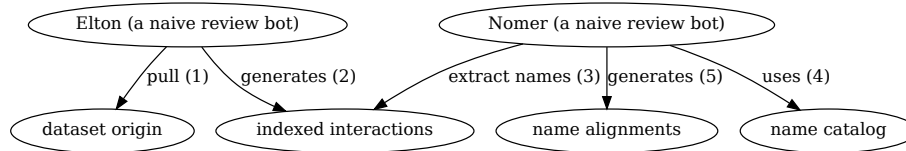
or visually, in a process diagram.

Figure 1: Review Process Overview

You can find a copy of the full review script at check-data.sh. See also GitHub and Codeberg.

Results

In the following sections, the results of the review are summarized [2]. Then, links to the detailed review reports are provided.

Files

The following files are produced in this review:

filename	description
biblio.bib	list of bibliographic reference of this review
check-dataset.sh	data review workflow/process as expressed in a bash script
data.zip	a versioned Preston (Elliott et al. 2025) archive of the data under review
HEAD	the digital signature of the data under review
index.docx	review in MS Word format
index.html	review in HTML format
index.md	review in Pandoc markdown format
index.pdf	review in PDF format
indexed-citations.csv.gz	list of distinct reference citations for reviewed species interaction claims in gzipped comma-separated values file format

---

[2]Disclaimer: The results in this review should be considered friendly, yet naive, notes from an unsophisticated robot. Please keep that in mind when considering the review results.

filename	description
indexed-citations.html.gz	list of distinct reference citations for reviewed species interactions claims in gzipped html file format
indexed-citations.tsv.gz	list of distinct reference citations for reviewed species interaction claims in gzipped tab-separated values format
indexed-interactions-col-family-col-family.svg	network diagram showing the taxon family to taxon family interaction claims in the dataset under review as interpreted by the Catalogue of Life via Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024)
indexed-interactions-col-kingdom-col-kingdom.svg	network diagram showing the taxon kingdom to taxon kingom interaction claims in the dataset under review as interpreted by the Catalogue of Life via Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024)
indexed-interactions.csv.gz	species interaction claims indexed from the dataset under review in gzipped comma-separated values format
indexed-interactions.html.gz	species interaction claims indexed from the dataset under review in gzipped html format
indexed-interactions.tsv.gz	species interaction claims indexed from the dataset under review in gzipped tab-separated values format
indexed-interactions-sample.csv	list of species interaction claims indexed from the dataset under review in gzipped comma-separated values format
indexed-interactions-sample.html	first 500 species interaction claims indexed from the dataset under review in html format
indexed-interactions-sample.tsv	first 500 species interaction claims indexed from the dataset under review in tab-separated values format
indexed-names.csv.gz	taxonomic names indexed from the dataset under review in gzipped comma-separated values format
indexed-names.html.gz	taxonomic names found in the dataset under review in gzipped html format

filename	description
indexed-names.tsv.gz	taxonomic names found in the dataset under review in gzipped tab-separated values format
indexed-names-resolved-col.csv.gz	taxonomic names found in the dataset under review aligned with the Catalogue of Life as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-col.html.gz	taxonomic names found in the dataset under review aligned with the Catalogue of Life as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-col.tsv.gz	taxonomic names found in the dataset under review aligned with the Catalogue of Life as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-resolved-discoverlife.csv.gz	taxonomic names found in the dataset under review aligned with Discover Life bee species checklist as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-discoverlife.html.gz	taxonomic names found in the dataset under review aligned with Discover Life bee species checklist as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-discoverlife.tsv.gz	taxonomic names found in the dataset under review aligned with Discover Life bee species checklist as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format

filename	description
indexed-names-resolved-gbif.csv.gz	taxonomic names found in the dataset under review aligned with GBIF Backbone Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-gbif.html.gz	taxonomic names found in the dataset under review aligned with GBIF Backbone Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-gbif.tsv.gz	taxonomic names found in the dataset under review aligned with GBIF Backbone Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-resolved-itis.csv.gz	taxonomic names found in the dataset under review aligned with Integrated Taxonomic Information System (ITIS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-itis.html.gz	taxonomic names found in the dataset under review aligned with Integrated Taxonomic Information System (ITIS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-itis.tsv.gz	taxonomic names found in the dataset under review aligned with Integrated Taxonomic Information System (ITIS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format

filename	description
indexed-names-resolved-mdd.csv.gz	taxonomic names found in the dataset under review aligned with the Mammal Diversity Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-mdd.html.gz	taxonomic names found in the dataset under review aligned with Mammal Diversity Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-mdd.tsv.gz	taxonomic names found in the dataset under review aligned with Mammal Diversity Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-resolved-ncbi.csv.gz	taxonomic names found in the dataset under review aligned with the NCBI Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-ncbi.html.gz	taxonomic names found in the dataset under review aligned with the NCBI Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-ncbi.tsv.gz	taxonomic names found in the dataset under review aligned with the NCBI Taxonomy as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format

filename	description
indexed-names-resolved-pbdb.csv.gz	taxonomic names found in the dataset under review aligned with the Paleobiology Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-pbdb.html.gz	taxonomic names found in the dataset under review aligned with Paleobiology Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-pbdb.tsv.gz	taxonomic names found in the dataset under review aligned with Paleobiology Database as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-resolved-tpt.csv.gz	taxonomic names found in the dataset under review aligned with the Terrestrial Parasite Tracker (TPT) Taxonomic Resource as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-tpt.html.gz	taxonomic names found in the dataset under review aligned with the Terrestrial Parasite Tracker (TPT) Taxonomic Resource as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-tpt.tsv.gz	taxonomic names found in the dataset under review aligned with the Terrestrial Parasite Tracker (TPT) Taxonomic Resource as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format

filename	description
indexed-names-resolved-wfo.csv.gz	taxonomic names found in the dataset under review aligned with the World of Flora Online as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-wfo.html.gz	taxonomic names found in the dataset under review aligned with the World of Flora Online as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-wfo.tsv.gz	taxonomic names found in the dataset under review aligned with the World of Flora Online as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-resolved-worms.csv.gz	taxonomic names found in the dataset under review aligned with the World Register of Marine Species (WoRMS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped comma-separated values format
indexed-names-resolved-worms.html.gz	taxonomic names found in the dataset under review aligned with the World Register of Marine Species (WoRMS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped html format
indexed-names-resolved-worms.tsv.gz	taxonomic names found in the dataset under review aligned with the World Register of Marine Species (WoRMS) as accessed through the Nomer Corpus of Taxonomic Resources (J. H. (ed.). Poelen 2024) in gzipped tab-separated values format
indexed-names-sample.csv	first 500 taxonomic names found in the dataset under review in comma-separated values format
indexed-names-sample.html	first 500 taxonomic names found in the dataset under review in html format

filename	description
indexed-names-sample.tsv	first 500 taxonomic names found in the dataset under review in tab-separated values format
interaction.svg	diagram summarizing the data model used to index species interaction claims
nanopub-sample.trig	first 500 species interaction claims as expressed in the nanopub format (Kuhn and Dumontier 2014)
nanopub.trig.gz	species interaction claims as expressed in the nanopub format (Kuhn and Dumontier 2014)
process.svg	diagram summarizing the data review processing workflow
prov.nq	origin of the dataset under review as expressed in rdf/nquads
review.csv.gz	review notes associated with the dataset under review in gzipped comma-separated values format
review.html.gz	review notes associated with the dataset under review in gzipped html format
review.tsv.gz	review notes associated with the dataset under review in gzipped tab-separated values format
review-sample.csv	first 500 review notes associated with the dataset under review in comma-separated values format
review-sample.html	first 500 review notes associated with the dataset under review in html format
review-sample.tsv	first 500 review notes associated with the dataset under review in tab-separated values format
review.svg	a review badge generated as part of the dataset review process
zenodo.json	metadata of this review expressed in Zenodo record metadata

Archived Dataset

Note that *data.zip* file in this archive contains the complete, unmodified archived dataset under review.
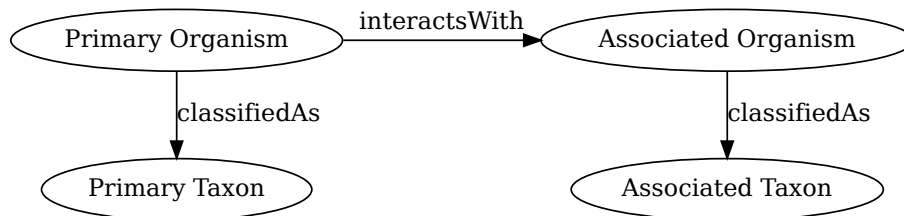
Biotic Interactions

Figure 2: Biotic Interaction Data Model

In this review, biotic interactions (or biotic associations) are modeled as a primary (aka subject, source) organism interacting with an associate (aka object, target) organism. The dataset under review classified the primary/associate organisms with specific taxa. The primary and associate organisms The kind of interaction is documented as an interaction type.

The dataset under review, named globalbioticinteractions/yale-peabody, has fingerprint hash://md5/41c7a250499cb3081abba26b82463ad9, is 466MiB in size and contains 283,454 interaction with 9 unique types of associations (e.g., adjacentTo) between 8,205 primary taxon (e.g., undet. Foraminiferida) and 140,237 associated taxon (e.g., rocks).

An exhaustive list of indexed interaction claims can be found in gzipped csv and tsv archives. To facilitate discovery, a preview of claims available in the gzipped html page at indexed-interactions.html.gz are shown below.

The exhaustive list was used to create the following data summaries below.

Table 3: Sample of Indexed Interaction Claims

sourceTaxonName	interactionTypeName	targetTaxonName	referenceCitation
Aegithalos caudatus macedonicus	interactsWith	ORN.113916	http://collections.peabody.yale.edu/search/Record ORN-065637
Erithacus rubecula rubecula	interactsWith	ORN.140511.001	http://collections.peabody.yale.edu/search/Record ORN-140507.001
Molothrus ater ater	hasHost	ORN.142308	http://collections.peabody.yale.edu/search/Record ORN-142309
Doricha enicura	interactsWith	ORN.142510.001	http://collections.peabody.yale.edu/search/Record ORN-142516

Table 4: Most Frequently Mentioned Interaction Types (up to 20 most frequent)

interactionTypeName	count
adjacentTo	153939
interactsWith	127289
hasHost	1761
hasParasite	386
symbiontOf	50
preysOn	13
preyedUponBy	7
visits	6
eats	3

Table 5: Most Frequently Mentioned Primary Taxa (up to 20 most frequent)

sourceTaxonName	count
undet. Foraminiferida	23302
Limopsis striatopunctatus	20122
undet. Bivalvia	16802
undet. Mollusca	13952
undet. Rugoglobigerinidae	5780
Protocardia sp.	4605
Hoploscaphites sp.	4431
Diptera	4400
undet. Rotaliida	4256
Protocardia subquadrata	4165
taxon undetermined	3972
Insecta	3617
Bivalvia	3357
undet. Conchostraca	3151
undet. Scaphitidae	2882
Hexapoda	2841
undet. Ammonoidea	2770
Coleoptera	2556
Mollusca	2317

Table 6: Most Frequently Mentioned Associate Taxa (up to 20 most frequent)

targetTaxonName	count
rocks	1435
same original lot: IP.004851	1158
same slab: IP.850994	774
same slab: IP.076166	713
same slide: IP.856874	704
same slide: IP.856866	654
same slide: IP.856867	635
trees	635
same slide: IP.856873	631
same slide: IP.856871	612
same slab: IP.351960	599
same slide: IP.856872	553
same slab: IP.392287	540
same slide: IP.856903	514
IP.856879	509
same slide: IP.856868	468
same slide: IP.856404	448
same slab: IP.903600	440
same slide: IP.856906	427

Table 7: Most Frequent Interactions between Primary and Associate Taxa (up to 20 most frequent)

sourceTaxonName	interactionTypeName	targetTaxonName	count
Limopsis striatopunctatus	adjacentTo	same slab: IP.850994	769
undet. Foraminiferida	adjacentTo	same slide: IP.856866	653
undet. Foraminiferida	adjacentTo	same slide: IP.856867	634
undet. Foraminiferida	adjacentTo	same slide: IP.856871	611
Corbulamella gregaria	adjacentTo	same slab: IP.076166	601
undet. Foraminiferida	adjacentTo	same slide: IP.856868	467
undet. Foraminiferida	interactsWith	IP.856879	411

sourceTaxonName	interactionTypeName	targetTaxonName	count
undet. Rotaliida	interactsWith	same original lot: IP.004851	389
undet. Foraminiferida	adjacentTo	same slide: IP.238722	389
undet. Foraminiferida	interactsWith	same original lot: IP.004851	371
Strophodonta sp.	interactsWith	same original lot: IP.558034	349
undet. Mollusca	adjacentTo	same slab: IP.316125	343
undet. Foraminiferida	adjacentTo	same slide: IP.856388	339
Protocardia subquadrata	adjacentTo	same slab: IP.392287	332
Protocardia sp.	adjacentTo	same slab: IP.391866	318
undet. Serpulidae	adjacentTo	same slab: IP.591512	317
Protocardia sp.	adjacentTo	same slab: IP.595261	298
undet. Foraminiferida	adjacentTo	same slide: IP.238721	297
Limopsis striatopunctatus	adjacentTo	same slab: IP.852383	295

Interaction Networks

The figures below provide a graph view on the dataset under review. The first shows a summary network on the kingdom level, and the second shows how interactions on the family level. It is important to note that both network graphs were first aligned taxonomically using the Catalogue of Life. Please refer to the original (or verbatim) taxonomic names for a more original view on the interaction data.

You can download the indexed dataset under review at indexed-interactions.csv .gz. A tab-separated file can be found at indexed-interactions.tsv.gz

Learn more about the structure of this download at GloBI website, by opening a GitHub issue, or by sending an email.

Another way to discover the dataset under review is by searching for it on the GloBI website.
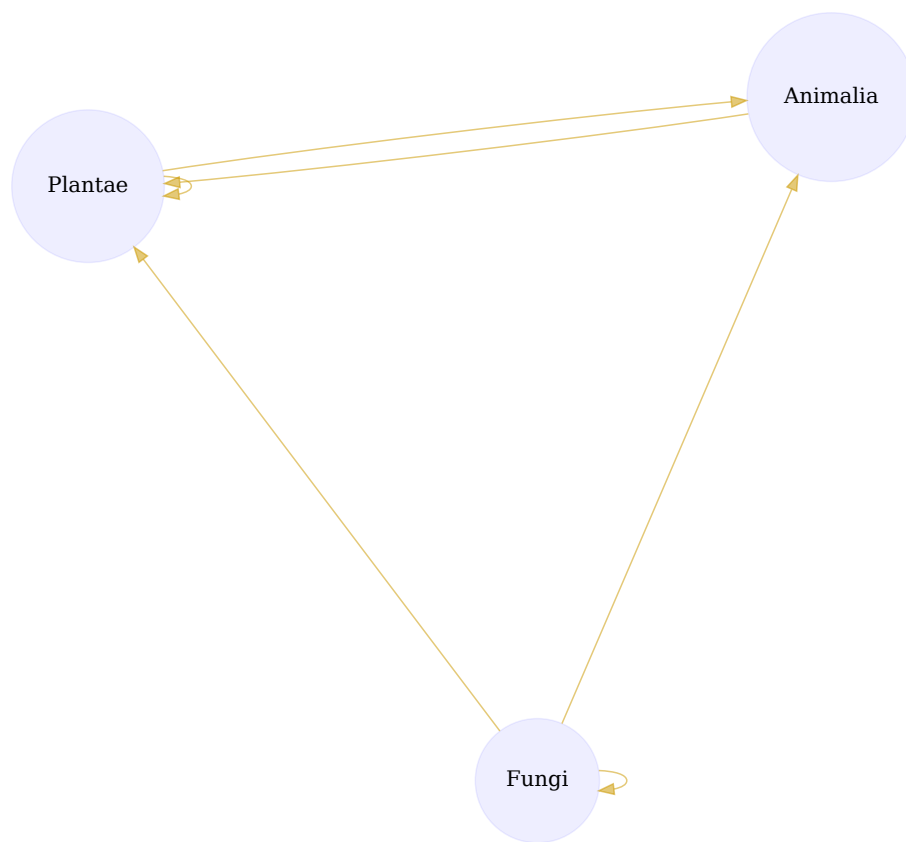
Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg
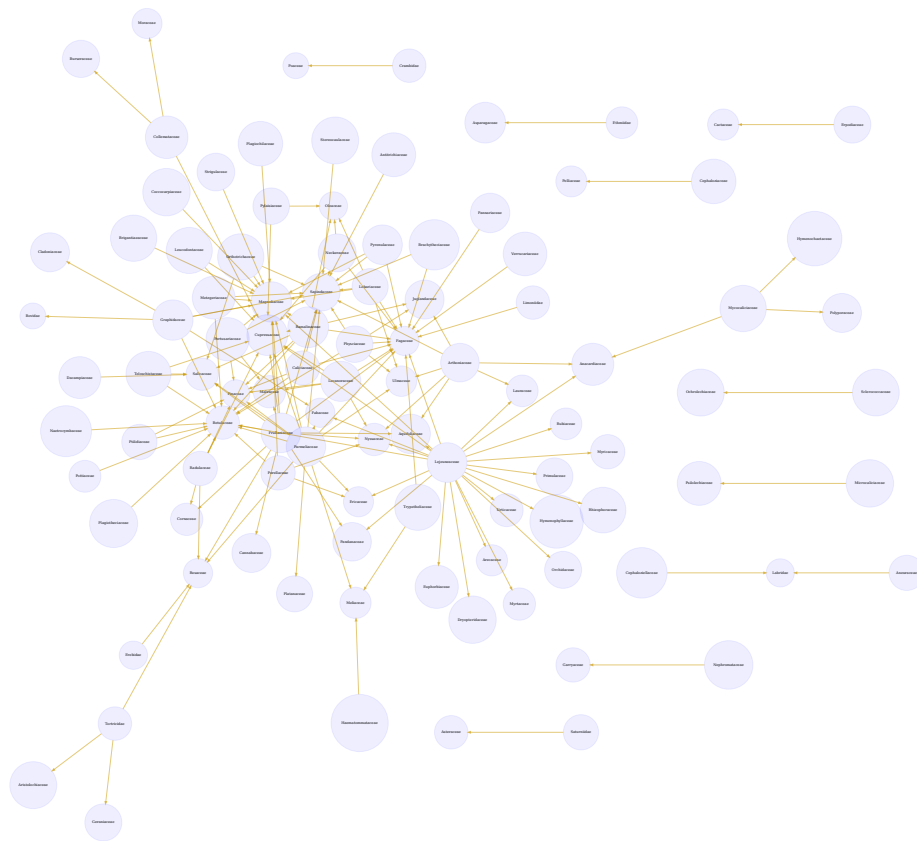
Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

Taxonomic Alignment

As part of the review, all names are aligned against various name catalogs (e.g., col, ncbi, discoverlife, gbif, itis, wfo, mdd, tpt, pbdb, and worms). These alignments can help review name usage or aid in selecting of a suitable taxonomic name resource.

Table 8: Sample of Name Alignments

providedName	relationName	resolvedCatalogName	resolvedName
Clay over sandstone rock at entrance triangular fault damp	NONE	col	Clay over sandstone rock at entrance triangular fault damp
Balsamea	SYNONYM_OF	col	Commiphora
Bank along an old road	NONE	col	Bank along an old road
Bank	NONE	col	Bank

Table 9: Distribution of Taxonomic Ranks of Aligned Names by Catalog. Names that were not aligned with a catalog are counted as NAs. So, the total number of unaligned names for a catalog will be listed in their NA row.

resolvedCatalogName	resolvedRank	count
col	NA	6732
col	class	37
col	family	199
col	genus	718
col	gigaclass	1
col	infraorder	5
col	infraphylum	1
col	infraspecific name	1
col	kingdom	3
col	order	64
col	parvorder	1
col	parvphylum	1
col	phylum	13
col	section	1
col	species	3351
col	subclass	10
col	subfamily	11
col	subgenus	20

18

resolvedCatalogName	resolvedRank	count
col	suborder	11
col	subphylum	4
col	subspecies	187
col	subtribe	1
col	superclass	1
col	superfamily	19
col	superorder	2
col	tribe	3
col	variety	22
discoverlife	NA	11352
discoverlife	species	13
gbif	NA	4891
gbif	class	37
gbif	family	280
gbif	form	9
gbif	genus	1294
gbif	kingdom	4
gbif	order	61
gbif	phylum	13
gbif	species	4559
gbif	subspecies	216
gbif	variety	74
itis	NA	8071
itis	class	38
itis	family	171
itis	genus	425
itis	infraclass	1
itis	infraorder	5
itis	infraphylum	1
itis	kingdom	3
itis	order	71
itis	phylum	12
itis	section	1
itis	species	2296
itis	subclass	16
itis	subfamily	9
itis	suborder	24
itis	subphylum	10
itis	subspecies	175
itis	superclass	2
itis	superfamily	11
itis	superorder	4
itis	variety	34
mdd	NA	11364

resolvedCatalogName	resolvedRank	count
ncbi	NA	8484
ncbi	clade	13
ncbi	class	35
ncbi	cohort	2
ncbi	family	165
ncbi	genus	452
ncbi	infraclass	1
ncbi	infraorder	6
ncbi	kingdom	2
ncbi	order	65
ncbi	phylum	13
ncbi	species	1999
ncbi	subclass	14
ncbi	subfamily	9
ncbi	subgenus	7
ncbi	suborder	15
ncbi	subphylum	4
ncbi	subspecies	68
ncbi	superclass	2
ncbi	superfamily	16
ncbi	superorder	5
ncbi	tribe	2
ncbi	varietas	1
pbdb	NA	8397
pbdb	class	60
pbdb	family	265
pbdb	genus	1073
pbdb	informal	1
pbdb	infraclass	2
pbdb	infraorder	7
pbdb	kingdom	5
pbdb	order	129
pbdb	phylum	16
pbdb	species	1305
pbdb	subclass	18
pbdb	subfamily	18
pbdb	suborder	39
pbdb	subphylum	7
pbdb	subspecies	16
pbdb	superclass	4
pbdb	superfamily	35
pbdb	superorder	5
pbdb	superphylum	1
pbdb	tribe	2

resolvedCatalogName	resolvedRank	count
pbdb	unranked clade	30
tpt	NA	11172
tpt	family	6
tpt	genus	14
tpt	order	5
tpt	species	167
wfo	NA	9517
wfo	class	1
wfo	family	7
wfo	genus	139
wfo	order	1
wfo	phylum	1
wfo	section	1
wfo	species	1667
wfo	subspecies	19
wfo	variety	17
worms	NA	9224
worms	class	46
worms	family	182
worms	genus	641
worms	gigaclass	1
worms	infraclass	5
worms	infraorder	6
worms	infraphylum	1
worms	kingdom	3
worms	order	71
worms	parvphylum	1
worms	phylum	15
worms	species	1086
worms	subclass	12
worms	subfamily	4
worms	suborder	22
worms	subphylum	9
worms	subspecies	16
worms	superclass	2
worms	superfamily	25
worms	superorder	2
worms	variety	1

Table 10: Name relationship types per catalog. Name relationship type "NONE" means that a name was not recognized by the associated catalog. "SAME_AS" indicates either a "HAS_ACCEPTED_NAME" or "SYNONYM_OF" name relationship type. We recognize that "SYNONYM_OF" encompasses many types of nomenclatural synonymies (ICZN 1999) (e.g., junior synonym, senior synonyms).

resolvedCatalogName	relationName	count
col	NONE	143265
col	SYNONYM_OF	2097
col	HAS_ACCEPTED_NAME	3891
discoverlife	NONE	148580
discoverlife	HAS_ACCEPTED_NAME	12
discoverlife	SYNONYM_OF	2
gbif	NONE	141223
gbif	SYNONYM_OF	3176
gbif	HAS_ACCEPTED_NAME	6367
itis	NONE	144820
itis	HAS_ACCEPTED_NAME	3048
itis	SYNONYM_OF	816
mdd	NONE	148471
mdd	HAS_ACCEPTED_NAME	102
mdd	SYNONYM_OF	1
ncbi	NONE	145171
ncbi	SAME_AS	3085
ncbi	SYNONYM_OF	377
pbdb	NONE	144726
pbdb	HAS_ACCEPTED_NAME	3655
pbdb	SYNONYM_OF	586
tpt	NONE	148367
tpt	HAS_ACCEPTED_NAME	202
tpt	SYNONYM_OF	6
wfo	NONE	146528
wfo	SYNONYM_OF	463
wfo	HAS_ACCEPTED_NAME	1333
wfo	HAS_UNCHECKED_NAME	368
worms	NONE	145894
worms	HAS_ACCEPTED_NAME	2432
worms	SYNONYM_OF	423

Table 11: List of Available Name Alignment Reports

catalog name	alignment results
col	associated names alignments report in gzipped html, csv, and tsv)
ncbi	associated names alignments report in gzipped html, csv, and tsv)
discoverlife	associated names alignments report in gzipped html, csv, and tsv)
gbif	associated names alignments report in gzipped html, csv, and tsv)
itis	associated names alignments report in gzipped html, csv, and tsv)
wfo	associated names alignments report in gzipped html, csv, and tsv)
mdd	associated names alignments report in gzipped html, csv, and tsv)
tpt	associated names alignments report in gzipped html, csv, and tsv)
pbdb	associated names alignments report in gzipped html, csv, and tsv)
worms	associated names alignments report in gzipped html, csv, and tsv)

Additional Reviews

Elton, Nomer, and other tools may have difficulties interpreting existing species interaction datasets. Or, they may misbehave, or otherwise show unexpected behavior. As part of the review process, detailed review notes are kept that document possibly misbehaving, or confused, review bots. An sample of review notes associated with this review can be found below.

Table 12: First few lines in the review notes.

reviewDate	reviewCommentType	reviewComment
2025-04-30T19:26:49Z	summary	https://github.com/globalbioticinteractions/yale-peabody/archive/4525d8b90998b1d3ec3c83a5074839fce5f
2025-04-30T19:26:49Z	summary	283454 interaction(s)
2025-04-30T19:26:49Z	summary	0 note(s)
2025-04-30T19:26:49Z	summary	20 info(s)

In addition, you can find the most frequently occurring notes in the table below.

: Most frequently occurring review notes, if any.

For additional information on review notes, please have a look at the first 500 Review Notes in html format or the download full gzipped csv or tsv archives.

GloBI Review Badge

As part of the review, a review badge is generated. This review badge can be included in webpages to indicate the review status of the dataset under review.

Figure 5: Picture of a GloBI Review Badge [3]

Note that if the badge is green, no review notes were generated. If the badge is yellow, the review bots may need some help with interpreting the species interaction data.

GloBI Index Badge

If the dataset under review has been registered with GloBI, and has been succesfully indexed by GloBI, the GloBI Index Status Badge will turn green. This means that the dataset under review was indexed by GloBI and is available through GloBI services and derived data products.

Figure 6: Picture of a GloBI Index Badge [4]

If you'd like to keep track of reviews or index status of the dataset under review, please visit GloBI's dataset index [5] for badge examples.

Discussion

This review and archive provides a means of creating citable versions of datasets that change frequently. This may be useful for dataset managers, including natural history collection data managers, as a backup archive of a shared Darwin Core archive. It also serves as a means of creating a trackable citation for the dataset in an automated way, while also including some information about the contents of the dataset.

This review aims to provide a perspective on the dataset to aid in understanding of species interaction claims discovered. However, it is important to note that

---

[3] Up-to-date status of the GloBI Review Badge can be retrieved from the GloBI Review Depot

[4] Up-to-date status of the GloBI Index Badge can be retrieved from GloBI's API

[5] At time of writing (2025-04-30) the version of the GloBI dataset index was available at https://globalbioticinteractions.org/datasets

this review does *not* assess the quality of the dataset. Instead, it serves as an indication of the open-ness[6] and FAIRness (Wilkinson et al. 2016; Trekels et al. 2023) of the dataset: to perform this review, the data was likely openly available, **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. The current Open-FAIR assessment is qualitative, and a more quantitative approach can be implemented with specified measurement units.

This report also showcases the reuse of machine-actionable (meta)data, something highly recommended by the FAIR Data Principles (Wilkinson et al. 2016). Making (meta)data machine-actionable enables more precise procesing by computers, enabling even naive review bots like Nomer and Elton to interpret the data effectively. This capability is crucial for not just automating the generation of reports, but also for facilitating seamless data exchanges, promoting interoperability.

Acknowledgements

Author contributions

Nomer was responsible for name alignments. Elton carried out dataset extraction, and generated the review notes. Preston tracked, versioned, and packaged, the dataset under review.

References

Elliott, Michael, Jorrit Poelen, Icaro Alzuru, Emilio Berti, and partha04patel. 2025. "Bio-Guoda/Preston: 0.10.5." Zenodo. https://doi.org/10.5281/zenodo.14662206.

ICZN. 1999. "International Code of Zoological Nomenclature." The International Trust for Zoological Nomenclature, London, UK. https://www.iczn.org/the-code/the-code-online/.

Kuhn, Tobias, and Michel Dumontier. 2014. "Trusty URIs: Verifiable, Immutable, and Permanent Digital Artifacts for Linked Data." In *The Semantic Web: Trends and Challenges*, edited by Valentina Presutti, Claudia d'Amato,

---

[6]According to http://opendefinition.org/: "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."

Fabien Gandon, Mathieu d'Aquin, Steffen Staab, and Anna Tordai, 395–410. Cham: Springer International Publishing.

Kuhn, Tobias, Jorrit Poelen, and Katrin Leinweber. 2025. "Globalbioticinteractions/Elton: 0.15.1." Zenodo. https://doi.org/10.5281/zenodo.14927734.

Poelen, Jorrit H. (ed.). 2024. "Nomer Corpus of Taxonomic Resources Hash://Sha256/ B60c0d25a16ae77b24305782017b1a270b79b5d1746f832650 F2027ba536e276 Hash://Md5/17f1363a277ee0e4ecaf1b91c665e47e." Zenodo. https://doi.org/10.5281/zenodo.12695629.

Poelen, Jorrit H., James D. Simons, and Chris J. Mungall. 2014. "Global Biotic Interactions: An Open Infrastructure to Share and Analyze Species-Interaction Datasets." *Ecological Informatics* 24 (November): 148–59. https://doi.org/10.1016/j.ecoinf.2014.08.005.

Poelen, Jorrit, Katja Seltmann, and Daniel Mietchen. 2024. "Globalbioticinteractions/Globinizer: 0.4.0." Zenodo. https://doi.org/10.5281/zenodo.10647565.

Salim, José Augusto, and Jorrit Poelen. 2025. "Globalbioticinteractions/Nomer: 0.5.15." Zenodo. https://doi.org/10.5281/zenodo.14893840.

Trekels, Maarten, Debora Pignatari Drucker, José Augusto Salim, Jeff Ollerton, Jorrit Poelen, Filipi Miranda Soares, Max Rünzel, Muo Kasina, Quentin Groom, and Mariano Devoto. 2023. "WorldFAIR Project (D10.1) Agriculture-related pollinator data standards use cases report." Zenodo. https://doi.org/10.5281/zenodo.8176978.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.