A Review of Biotic Interactions and Taxon Names Found in globalbioticinteractions/unl-nsm hash://md5/80879a196ada83aecd073df613018232

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

2025-04-22

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/unl-nsm, has fingerprint hash://md5/80879a196ada83aecd073df613018232, is 47.4MiB in size and contains 106,383 interaction with 3 unique types of associations (e.g., hasHost) between 5,804 primary taxon (e.g., Arthropoda) and 6,465 associated taxon (e.g., Microtus oeconomus). This report includes detailed summaries of interaction data, a taxonomic review from multiple catalogs, and an archived version of the dataset from which the reviews are derived.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

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:

> Harold W. Manter Laboratory of Parasitology Collection (HWML)
> Parasite Collection (Arctos) - Version 1.8 http://ipt.vertnet.org:8080/ipt/archive.do?r=hwml_para
> 2025-04-19T03:54:26.501Z hash://md5/80879a196ada83aecd073df613018232

For additional metadata related to this dataset, please visit https://github.com/globalbioticinteractions/unl-nsm 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

tool name	version
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. 47.4MiB) under review
elton pull globalbioticinteractions/unl-nsm

# generate review notes
elton review globalbioticinteractions/unl-nsm\
 > review.tsv

# export indexed interaction records
elton interactions globalbioticinteractions/unl-nsm\
 > interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names globalbioticinteractions/unl-nsm\
 | nomer append col\
 > name-alignment.tsv
```
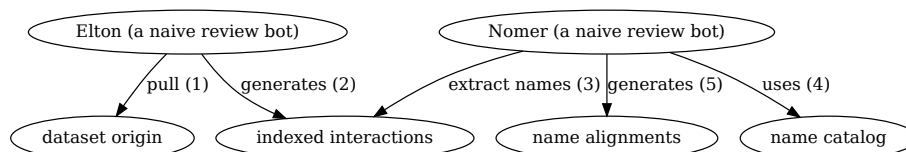
or visually, in a process diagram.

Figure 1: Review Process Overview

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

---

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

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
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)

---

[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-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
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

filename	description
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
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

filename	description
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
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

filename	description
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
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

filename	description
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
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

filename	description
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
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

filename	description
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

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
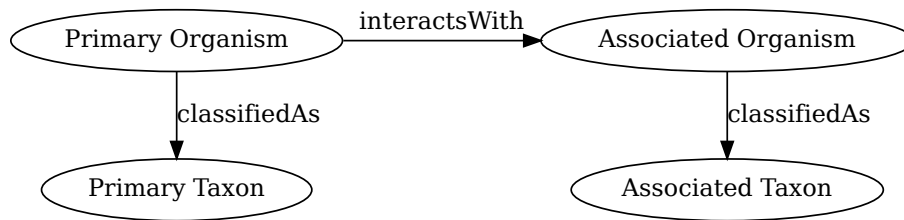
11

Figure 2: Biotic Interaction Data Model

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/unl-nsm, has fingerprint hash://md5/80879a196ada83aecd073df613018232, is 47.4MiB in size and contains 106,383 interaction with 3 unique types of associations (e.g., hasHost) between 5,804 primary taxon (e.g., Arthropoda) and 6,465 associated taxon (e.g., Microtus oeconomus).

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
Plasmodium attenuatum	hasHost	Ameiva ameiva	http://arctos.database.museum/guid/HWML:Par
Ornithodoros	hasHost	Tomopeas ravus	http://arctos.database.museum/guid/HWML:Par
Analges	hasHost	Corvus brachyrhynchos	http://arctos.database.museum/guid/HWML:Par
Arthropoda	hasHost	Porzana carolina	http://arctos.database.museum/guid/HWML:Par

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

interactionTypeName	count
hasHost	105465
parasiteOf	914
coOccursWith	4

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

sourceTaxonName	count
Arthropoda	5176
Amalaraeus dissimilis	2914
Pulex irritans	2866
Orchopeas caedens	2369
Peromyscopsylla ostsibirica	1941
Euhoplopsyllus glacialis lynx	1858
Orchopeas sexdentatus	1748
Megabothris quirini	1709
Monopsyllus vison	1615
Orchopeas leucopus	1553
Megabothris calcarifer	1551
Hepatozoon	1541
Ceratophyllus riparius	1515
Monopsyllus wagneri	1426
Catallagia dacenkoi	1417
Thrassis bacchi	1356
Opisocrostis hirsutus	1215
Cestoda	1207
Acari	1173

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

targetTaxonName	count
Microtus oeconomus	5084
Clethrionomys rutilus	4769
Peromyscus maniculatus	3732
Tamiasciurus hudsonicus	3199
Canis latrans	2270
Procyon lotor	1731
Riparia riparia	1539
Peromyscus leucopus	1343
Ammospermophilus leucurus	1319
Canis familiaris	1166
Lynx canadensis	1100
Petrochelidon pyrrhonota	994
Peromyscus	973
Cynomys ludovicianus	960
Takydromus tachydromoides	958

targetTaxonName	count
Microtus pennsylvanicus	945
Martes americana	820
Homo sapiens	777
Sylvilagus floridanus	733

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

sourceTaxonName	interactionTypeName	targetTaxonName	count
Amalaraeus dissimilis	hasHost	Clethrionomys rutilus	1641
Monopsyllus vison	hasHost	Tamiasciurus hudsonicus	1484
Peromyscopsylla ostsibirica	hasHost	Microtus oeconomus	1476
Orchopeas caedens	hasHost	Tamiasciurus hudsonicus	1414
Pulex irritans	hasHost	Canis latrans	1301
Ceratophyllus riparius	hasHost	Riparia riparia	1194
Megabothris calcarifer	hasHost	Microtus oeconomus	1143
Thrassis bacchi	hasHost	Ammospermophilus leucurus	1118
Monopsyllus wagneri	hasHost	Peromyscus maniculatus	1059
Megabothris quirini	hasHost	Clethrionomys rutilus	1044
Euhoplopsyllus glacialis lynx	hasHost	Lynx canadensis	986
Ceratophyllus scopulorum	hasHost	Petrochelidon pyrrhonota	861
Amalaraeus dissimilis	hasHost	Microtus oeconomus	799
Catallagia dacenkoi	hasHost	Clethrionomys rutilus	797
Opisocrostis hirsutus	hasHost	Cynomys ludovicianus	695
Orchopeas caedens	hasHost	Martes americana	545
Megabothris quirini	hasHost	Microtus oeconomus	527
Oropsylla idahoensis	hasHost	Citellus undulatus	491
Catallagia dacenkoi	hasHost	Microtus oeconomus	469

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.
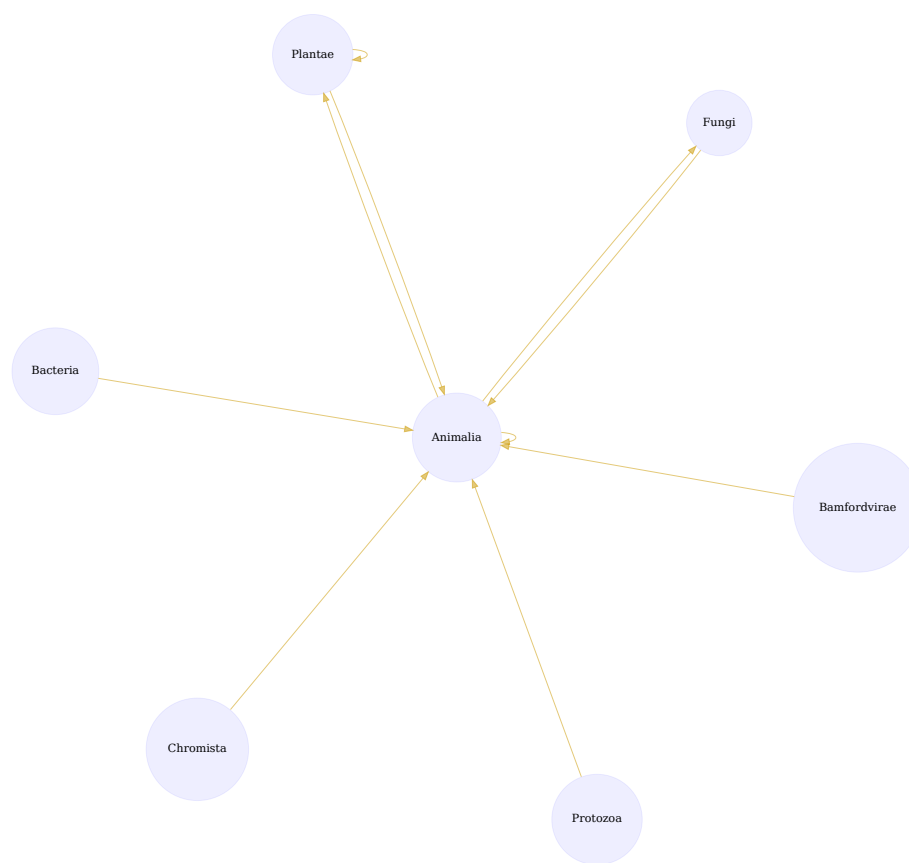
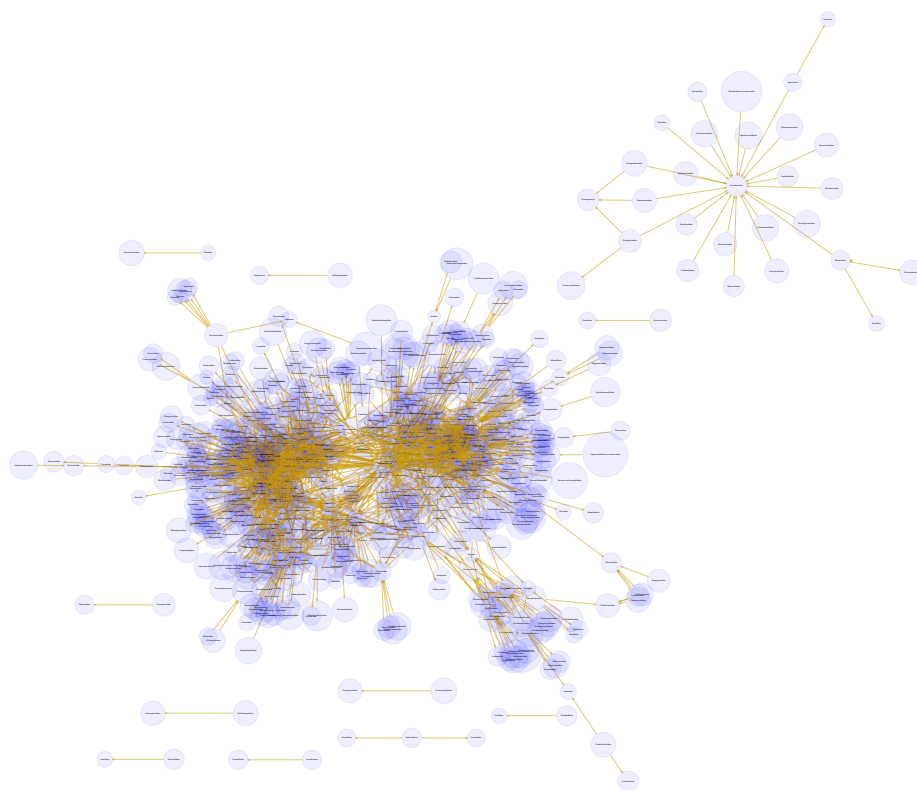Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

Another way to discover the dataset under review is by searching for it on the GloBI website.

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
Aves	HAS_ACCEPTED_NAME	col	Aves
Canis familiaris	SYNONYM_OF	col	Canis lupus familiaris
Catharus ustuslatus	NONE	col	Catharus ustuslatus
Coccyzus americanus	HAS_ACCEPTED_NAME	col	Coccyzus americanus

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	4078
col	class	12
col	family	99
col	form	1
col	genus	1079
col	gigaclass	1
col	infraorder	1
col	infraspecific name	2
col	kingdom	2
col	nanorder	1
col	order	19
col	parvorder	1
col	parvphylum	1
col	phylum	7
col	species	5970
col	subclass	3
col	subfamily	4
col	subgenus	38
col	suborder	2
col	subspecies	149

17

resolvedCatalogName	resolvedRank	count
col	superfamily	6
col	superorder	1
col	tribe	1
col	unranked	1
discoverlife	NA	11433
discoverlife	species	1
gbif	NA	3142
gbif	class	14
gbif	family	104
gbif	form	2
gbif	genus	1268
gbif	kingdom	3
gbif	order	19
gbif	phylum	7
gbif	species	6717
gbif	subspecies	174
itis	NA	6166
itis	class	11
itis	family	95
itis	genus	907
itis	infraorder	1
itis	infraphylum	1
itis	kingdom	2
itis	order	22
itis	phylum	8
itis	species	4109
itis	subclass	8
itis	subfamily	4
itis	subgenus	4
itis	suborder	4
itis	subphylum	1
itis	subspecies	93
itis	superclass	1
itis	superfamily	2
itis	superorder	1
mdd	NA	11433
ncbi	NA	5962
ncbi	clade	2
ncbi	class	11
ncbi	family	98
ncbi	genus	1046
ncbi	infraorder	1
ncbi	kingdom	1
ncbi	order	21

resolvedCatalogName	resolvedRank	count
ncbi	phylum	7
ncbi	species	4198
ncbi	subclass	8
ncbi	subfamily	4
ncbi	subgenus	11
ncbi	suborder	3
ncbi	subphylum	1
ncbi	subspecies	62
ncbi	superclass	1
ncbi	superfamily	7
ncbi	superkingdom	1
ncbi	superorder	2
pbdb	NA	9589
pbdb	class	14
pbdb	family	43
pbdb	genus	417
pbdb	infraorder	1
pbdb	kingdom	2
pbdb	order	16
pbdb	phylum	5
pbdb	species	1323
pbdb	subclass	1
pbdb	subfamily	3
pbdb	subgenus	1
pbdb	subkingdom	1
pbdb	suborder	6
pbdb	subspecies	2
pbdb	superclass	1
pbdb	superfamily	4
pbdb	superorder	2
pbdb	superphylum	1
pbdb	tribe	1
pbdb	unranked clade	14
tpt	NA	9348
tpt	family	24
tpt	genus	226
tpt	order	4
tpt	species	1831
wfo	NA	11387
wfo	genus	38
wfo	species	8
worms	NA	6049
worms	class	11
worms	family	84

resolvedCatalogName	resolvedRank	count
worms	forma	1
worms	genus	910
worms	gigaclass	1
worms	infraorder	1
worms	infraphylum	1
worms	kingdom	3
worms	order	22
worms	parvphylum	1
worms	phylum	6
worms	phylum (division)	1
worms	species	4310
worms	subclass	7
worms	subfamily	4
worms	subgenus	3
worms	suborder	4
worms	subphylum	2
worms	subspecies	7
worms	superclass	1
worms	superfamily	7
worms	superorder	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	HAS_ACCEPTED_NAME	6297
col	SYNONYM_OF	2039
col	NONE	4780
discoverlife	NONE	12269
discoverlife	HAS_ACCEPTED_NAME	1
gbif	HAS_ACCEPTED_NAME	7023
gbif	SYNONYM_OF	2632
gbif	NONE	3843
itis	HAS_ACCEPTED_NAME	4544
itis	SYNONYM_OF	937
itis	NONE	6868
mdd	NONE	11594

resolvedCatalogName	relationName	count
mdd	HAS_ACCEPTED_NAME	630
mdd	SYNONYM_OF	38
ncbi	SAME_AS	4859
ncbi	SYNONYM_OF	757
ncbi	NONE	6677
ncbi	COMMON_NAME_OF	9
pbdb	HAS_ACCEPTED_NAME	1725
pbdb	SYNONYM_OF	298
pbdb	NONE	10310
tpt	NONE	10097
tpt	HAS_ACCEPTED_NAME	2233
tpt	SYNONYM_OF	370
wfo	NONE	12216
wfo	HAS_ACCEPTED_NAME	26
wfo	SYNONYM_OF	19
wfo	HAS_UNCHECKED_NAME	6
worms	HAS_ACCEPTED_NAME	4191
worms	NONE	6805
worms	SYNONYM_OF	1400

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)

catalog name	alignment results
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
2025-04-22T16:03:42Z	summary	https://github.com/globalbioticinteractions/unl-nsm/archive/6bcd8aec22e4309b7f4e8be1afe8191d391e73c
2025-04-22T16:03:42Z	summary	106383 interaction(s)
2025-04-22T16:03:42Z	summary	0 note(s)
2025-04-22T16:03:42Z	summary	2 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.

---

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

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 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.

---

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

[5]At time of writing (2025-04-22) the version of the GloBI dataset index was available at https://globalbioticinteractions.org/datasets

[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."

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, 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.