A Review of Biotic Interactions and Taxon Names Found in EMTuckerLabUMMZ/ummzi hash://md5/3a3f825c7671f3ff1d3acdf6675cfecb

by Nomer, Elton and Preston, three naive review bots

review@globalbioticinteractions.org

https://globalbioticinteractions.org/contribute

https://github.com/EMTuckerLabUMMZ/ummzi/issues

2025-04-10

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 EMTuckerLabUMMZ/ummzi, has fingerprint hash://md5/3a3f825c7671f3ff1d3acdf6675cfecb, is 211MiB in size and contains 94,489 interaction with 1 unique type of association (e.g., hasHost) between 766 primary taxa (e.g., Acari) and 12,548 associated taxon (e.g., Acrocephalus arundinaceus). This report includes detailed summaries of interaction data, a taxonomic review from multiple catalogs, and an archived version of the dataset from which the reviews are derived.

Table of 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:

University of Michigan Museum of Zoology Insect Division. Full Database Export 2020-11-20 provided by Erika Tucker and Barry Oconner. https://github.com/EMTuckerLabUMMZ/ummzi/archive/6731357a377e9c2748fc931faa2ff3dc0ce3ea7a.zip 2025-04-04T23:02:16.035Z hash://md5/3a3f825c7671f3ff1d3acdf6675cfecb

For additional metadata related to this dataset, please visit <https://github.com/EMTuckerLabUMMZ/ummzi> 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.

Tools used in this review process

| tool name | version |
| --- | --- |
| [preston](https://github.com/bio-guoda/preston) | 0.10.1 |
| [elton](https://github.com/globalbioticinteractions/elton) | 0.15.9 |
| [nomer](https://github.com/globalbioticinteractions/nomer) | 0.5.13 |
| [globinizer](https://github.com/globalbioticinteractions/globinizer) | 0.4.0 |
| [mlr](https://miller.readthedocs.io/en/6.8.0/) | 6.0.0 |
| [jq](https://jqlang.org/) | 1.6 |
| [yq](https://mikefarah.gitbook.io/yq) | 4.25.3 |
| [pandoc](https://pandoc.org/) | 3.1.6.1 |

The review process can be described in the form of the script below [[1]](#footnote-31).

# get versioned copy of the dataset (size approx. 211MiB) under review   
elton pull EMTuckerLabUMMZ/ummzi  
  
# generate review notes  
elton review EMTuckerLabUMMZ/ummzi\  
 > review.tsv  
  
# export indexed interaction records  
elton interactions EMTuckerLabUMMZ/ummzi\  
 > interactions.tsv  
  
# export names and align them with the Catalogue of Life using Nomer   
elton names EMTuckerLabUMMZ/ummzi\  
 | nomer append col\  
 > name-alignment.tsv

or visually, in a process diagram.



Review Process Overview

You can find a copy of the full review script at [check-data.sh](check-dataset.sh). See also [GitHub](https://github.com/globalbioticinteractions/globinizer/blob/master/check-dataset.sh) and [Codeberg](https://codeberg.org/globalbioticinteractions/globinizer/src/branch/master/check-dataset.sh).

# Results

In the following sections, the results of the review are summarized [[2]](#footnote-40). 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) |
| <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 |
| <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 |
| <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 |
| <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 |
| <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 |
| <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 |
| <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*](data.zip) file in this archive contains the complete, unmodified archived dataset under review.

## Biotic Interactions



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 EMTuckerLabUMMZ/ummzi, has fingerprint hash://md5/3a3f825c7671f3ff1d3acdf6675cfecb, is 211MiB in size and contains 94,489 interaction with 1 unique type of association (e.g., hasHost) between 766 primary taxa (e.g., Acari) and 12,548 associated taxon (e.g., Acrocephalus arundinaceus).

An exhaustive list of indexed interaction claims can be found in gzipped [csv](indexed-interactions.csv.gz) and [tsv](indexed-interactions.tsv.gz) 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.

Sample of Indexed Interaction Claims

| sourceTaxonName | interactionTypeName | targetTaxonName | referenceCitation |
| --- | --- | --- | --- |
| Acari | hasHost | Scarabaeidae | https://doi.org/10.15468/tmxd7n |
| Acari | hasHost | Dermestes maculatus | https://doi.org/10.15468/tmxd7n |
| Acari | hasHost | Anthidium undulatiforme | https://doi.org/10.15468/tmxd7n |
| Acari | hasHost | Anthidium flavomaculatum | https://doi.org/10.15468/tmxd7n |

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

| interactionTypeName | count |
| --- | --- |
| hasHost | 94489 |

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

| sourceTaxonName | count |
| --- | --- |
| Acari | 57444 |
| Arthropoda | 9457 |
| Proctophyllodes | 3237 |
| Proterothrix | 1004 |
| Proctophyllodes polyxenus | 944 |
| Pterodectes | 645 |
| Bombus | 634 |
| Amerodectes | 584 |
| Trouessartia trouessarti | 558 |
| Hymenoptera | 487 |
| Proctophyllodes spini | 478 |
| Trouessartia lonchurae | 406 |
| Trouessartia trouessartii | 363 |
| Ceratina | 355 |
| Nycteridocaulus | 354 |
| Proctophyllodes cotyledon | 304 |
| Apis | 293 |
| Proctophyllodes egglestoni | 273 |
| Proctophyllodes hylocichlae | 249 |

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

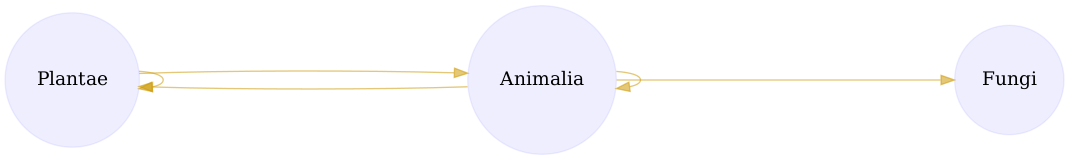
| targetTaxonName | count |
| --- | --- |
| Acrocephalus arundinaceus | 1340 |
| Hirundo rustica | 1101 |
| Turdus migratorius | 494 |
| Spinus tristis | 448 |
| Lophuromys aquilus | 424 |
| Myotis velifer | 416 |
| Emberiza rutila | 375 |
| Geospiza fuliginosa | 368 |
| Praomys taitae | 350 |
| Quiscalus quiscula | 350 |
| Bombycilla cedrorum | 336 |
| Haemorhous mexicanus | 334 |
| Pheucticus ludovicianus | 330 |
| Seiurus aurocapilla | 294 |
| Solidago | 256 |
| Centaurea stoebe | 256 |
| Lonchura leucogastra | 252 |
| Junco hyemalis | 249 |
| Chilocorus cacti | 239 |

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

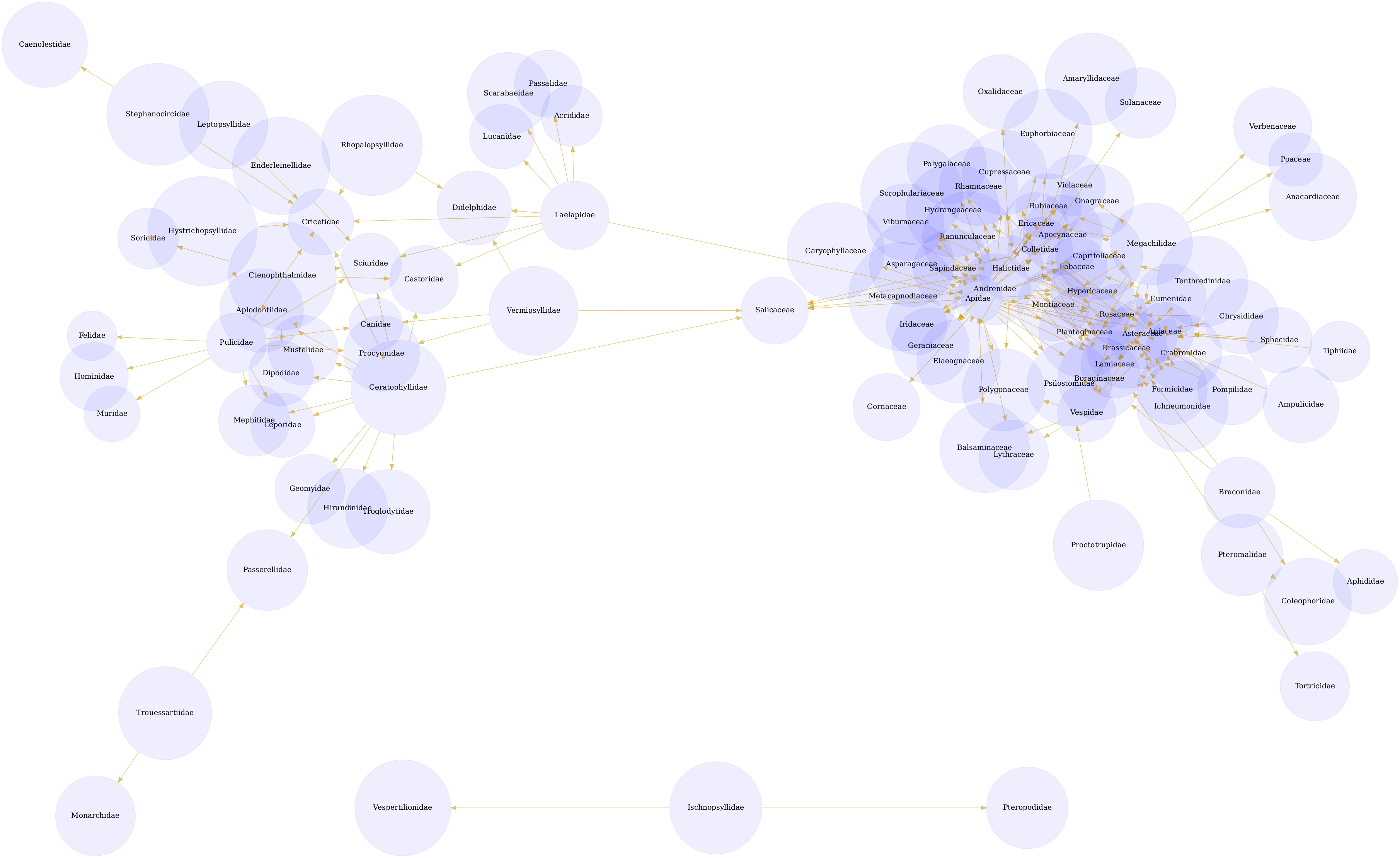
| sourceTaxonName | interactionTypeName | targetTaxonName | count |
| --- | --- | --- | --- |
| Acari | hasHost | Hirundo rustica | 867 |
| Trouessartia trouessarti | hasHost | Acrocephalus arundinaceus | 515 |
| Acari | hasHost | Acrocephalus arundinaceus | 421 |
| Acari | hasHost | Myotis velifer | 416 |
| Acari | hasHost | Lophuromys aquilus | 387 |
| Acari | hasHost | Geospiza fuliginosa | 368 |
| Acari | hasHost | Praomys taitae | 350 |
| Trouessartia trouessartii | hasHost | Acrocephalus arundinaceus | 349 |
| Proctophyllodes | hasHost | Turdus migratorius | 336 |
| Proctophyllodes spini | hasHost | Spinus tristis | 308 |
| Acari | hasHost | Emberiza rutila | 288 |
| Acari | hasHost | Chilocorus cacti | 239 |
| Proctophyllodes ampelidis | hasHost | Bombycilla cedrorum | 234 |
| Proctophyllodes | hasHost | Haemorhous mexicanus | 204 |
| Acari | hasHost | Microgale cowani | 169 |
| Acari | hasHost | Zosterops erythropleurus | 166 |
| Proctophyllodes polyxenus | hasHost | Junco hyemalis | 162 |
| Acari | hasHost | Microtus pennsylvanicus | 159 |
| Acari | hasHost | Turdus migratorius | 158 |

### 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.



Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life [download svg](indexed-interactions-col-kingdom-col-kingdom.svg)



Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. [download svg](indexed-interactions-col-family-col-family.svg)

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](https://globalbioticinteractions.org), by opening a [GitHub issue](https://github.com/globalbioticinteractions/globalbioticinteractions/issues/new), or by sending an [email](mailto:info@globalbioticinteractions.org).

Another way to discover the dataset under review is by searching for it on the [GloBI website](https://www.globalbioticinteractions.org/?accordingTo=globi%3AEMTuckerLabUMMZ%2Fummzi).

## 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.

Sample of Name Alignments

| providedName | relationName | resolvedCatalogName | resolvedName |
| --- | --- | --- | --- |
| Reserved by for trp | NONE | col | Reserved by for trp |
| Abditomys latidens | HAS\_ACCEPTED\_NAME | col | Abditomys latidens |
| Abeillia abeillei | HAS\_ACCEPTED\_NAME | col | Abeillia abeillei |
| Abies alba | HAS\_ACCEPTED\_NAME | col | Abies alba |

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 | 2763 |
| col | class | 6 |
| col | family | 150 |
| col | genus | 543 |
| col | infraorder | 1 |
| col | order | 23 |
| col | phylum | 3 |
| col | species | 7764 |
| col | subfamily | 2 |
| col | subgenus | 29 |
| col | subspecies | 1381 |
| col | superfamily | 1 |
| col | tribe | 3 |
| col | variety | 8 |
| discoverlife | NA | 11629 |
| discoverlife | species | 983 |
| gbif | NA | 2109 |
| gbif | class | 6 |
| gbif | family | 157 |
| gbif | form | 11 |
| gbif | genus | 636 |
| gbif | order | 23 |
| gbif | phylum | 3 |
| gbif | species | 8260 |
| gbif | subspecies | 1457 |
| gbif | variety | 18 |
| itis | NA | 3481 |
| itis | class | 5 |
| itis | division | 2 |
| itis | family | 149 |
| itis | genus | 468 |
| itis | order | 25 |
| itis | phylum | 1 |
| itis | species | 7143 |
| itis | subclass | 1 |
| itis | subfamily | 3 |
| itis | subspecies | 1327 |
| itis | superfamily | 2 |
| itis | superorder | 1 |
| itis | tribe | 2 |
| itis | variety | 4 |
| mdd | NA | 12611 |
| ncbi | NA | 4448 |
| ncbi | class | 5 |
| ncbi | cohort | 1 |
| ncbi | family | 146 |
| ncbi | genus | 562 |
| ncbi | order | 25 |
| ncbi | phylum | 3 |
| ncbi | species | 6980 |
| ncbi | subclass | 1 |
| ncbi | subfamily | 6 |
| ncbi | subgenus | 28 |
| ncbi | subspecies | 428 |
| ncbi | superfamily | 1 |
| ncbi | tribe | 2 |
| ncbi | varietas | 2 |
| pbdb | NA | 10439 |
| pbdb | class | 7 |
| pbdb | family | 141 |
| pbdb | genus | 316 |
| pbdb | infraclass | 1 |
| pbdb | infraorder | 1 |
| pbdb | order | 27 |
| pbdb | phylum | 1 |
| pbdb | species | 1667 |
| pbdb | subclass | 1 |
| pbdb | subfamily | 5 |
| pbdb | suborder | 2 |
| pbdb | subspecies | 7 |
| pbdb | superfamily | 1 |
| pbdb | tribe | 2 |
| pbdb | unranked clade | 4 |
| tpt | NA | 7455 |
| tpt | family | 24 |
| tpt | genus | 162 |
| tpt | order | 2 |
| tpt | species | 4968 |
| wfo | NA | 12111 |
| wfo | family | 16 |
| wfo | genus | 109 |
| wfo | species | 369 |
| wfo | subspecies | 12 |
| wfo | variety | 4 |
| worms | NA | 10891 |
| worms | class | 5 |
| worms | family | 117 |
| worms | genus | 215 |
| worms | infraorder | 1 |
| worms | order | 24 |
| worms | phylum | 1 |
| worms | phylum (division) | 2 |
| worms | species | 1306 |
| worms | subclass | 1 |
| worms | subspecies | 53 |

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 | 2958 |
| col | HAS\_ACCEPTED\_NAME | 11019 |
| col | SYNONYM\_OF | 1871 |
| discoverlife | NONE | 14321 |
| discoverlife | SYNONYM\_OF | 303 |
| discoverlife | HAS\_ACCEPTED\_NAME | 939 |
| discoverlife | HOMONYM\_OF | 56 |
| gbif | NONE | 2253 |
| gbif | HAS\_ACCEPTED\_NAME | 12319 |
| gbif | SYNONYM\_OF | 2716 |
| itis | NONE | 3736 |
| itis | HAS\_ACCEPTED\_NAME | 10256 |
| itis | SYNONYM\_OF | 1449 |
| mdd | NONE | 14230 |
| mdd | HAS\_ACCEPTED\_NAME | 1085 |
| mdd | SYNONYM\_OF | 40 |
| ncbi | NONE | 4903 |
| ncbi | SAME\_AS | 9532 |
| ncbi | SYNONYM\_OF | 1083 |
| pbdb | NONE | 12151 |
| pbdb | HAS\_ACCEPTED\_NAME | 3015 |
| pbdb | SYNONYM\_OF | 244 |
| tpt | NONE | 8836 |
| tpt | HAS\_ACCEPTED\_NAME | 6279 |
| tpt | SYNONYM\_OF | 314 |
| wfo | NONE | 14777 |
| wfo | HAS\_ACCEPTED\_NAME | 525 |
| wfo | SYNONYM\_OF | 116 |
| wfo | HAS\_UNCHECKED\_NAME | 105 |
| worms | NONE | 12746 |
| worms | HAS\_ACCEPTED\_NAME | 2436 |
| worms | SYNONYM\_OF | 269 |

List of Available Name Alignment Reports

| catalog name | alignment results |
| --- | --- |
| col | associated names alignments report in gzipped [html](indexed-names-resolved-col.html.gz), [csv](indexed-names-resolved-col.csv.gz), and [tsv](indexed-names-resolved-col.tsv.gz)) |
| ncbi | associated names alignments report in gzipped [html](indexed-names-resolved-ncbi.html.gz), [csv](indexed-names-resolved-ncbi.csv.gz), and [tsv](indexed-names-resolved-ncbi.tsv.gz)) |
| discoverlife | associated names alignments report in gzipped [html](indexed-names-resolved-discoverlife.html.gz), [csv](indexed-names-resolved-discoverlife.csv.gz), and [tsv](indexed-names-resolved-discoverlife.tsv.gz)) |
| gbif | associated names alignments report in gzipped [html](indexed-names-resolved-gbif.html.gz), [csv](indexed-names-resolved-gbif.csv.gz), and [tsv](indexed-names-resolved-gbif.tsv.gz)) |
| itis | associated names alignments report in gzipped [html](indexed-names-resolved-itis.html.gz), [csv](indexed-names-resolved-itis.csv.gz), and [tsv](indexed-names-resolved-itis.tsv.gz)) |
| wfo | associated names alignments report in gzipped [html](indexed-names-resolved-wfo.html.gz), [csv](indexed-names-resolved-wfo.csv.gz), and [tsv](indexed-names-resolved-wfo.tsv.gz)) |
| mdd | associated names alignments report in gzipped [html](indexed-names-resolved-mdd.html.gz), [csv](indexed-names-resolved-mdd.csv.gz), and [tsv](indexed-names-resolved-mdd.tsv.gz)) |
| tpt | associated names alignments report in gzipped [html](indexed-names-resolved-tpt.html.gz), [csv](indexed-names-resolved-tpt.csv.gz), and [tsv](indexed-names-resolved-tpt.tsv.gz)) |
| pbdb | associated names alignments report in gzipped [html](indexed-names-resolved-pbdb.html.gz), [csv](indexed-names-resolved-pbdb.csv.gz), and [tsv](indexed-names-resolved-pbdb.tsv.gz)) |
| worms | associated names alignments report in gzipped [html](indexed-names-resolved-worms.html.gz), [csv](indexed-names-resolved-worms.csv.gz), and [tsv](indexed-names-resolved-worms.tsv.gz)) |

## 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.

First few lines in the review notes.

| reviewDate | reviewCommentType | reviewComment |
| --- | --- | --- |
| 2025-04-10T02:37:17Z | note | target taxon name missing |
| 2025-04-10T02:37:17Z | note | target taxon name missing |
| 2025-04-10T02:37:17Z | note | target taxon name missing |
| 2025-04-10T02:37:17Z | note | target taxon name missing |

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

Most frequently occurring review notes, if any.

| reviewComment | count |
| --- | --- |
| target taxon name missing | 367307 |
| target taxon name missing: using institutionCode/collectionCode/collectionId/catalogNumber/occurrenceId as placeholder | 255 |
| source taxon name missing: using institutionCode/collectionCode/collectionId/catalogNumber/occurrenceId as placeholder | 29 |

For additional information on review notes, please have a look at the first 500 [Review Notes](review-sample.html) in html format or the download full gzipped [csv](review.csv.gz) or [tsv](review.tsv.gz) 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.

Picture of a GloBI Review Badge 

Picture of a GloBI Review Badge [[3]](#footnote-134)

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](https://globalbioticinteractions.org/contribute), 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.

Picture of a GloBI Index Badge [[4]](#footnote-141)

If you’d like to keep track of reviews or index status of the dataset under review, please visit GloBI’s dataset index [[5]](#footnote-143) 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]](#footnote-147) 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

We thank the many humans that created us and those who created and maintained the data, software and other intellectual resources that were used for producing this review. In addition, we are grateful for the natural resources providing the basis for these human and bot activities. Also, thanks to https://github.com/zygoballus for helping improve the layout of the review tables.

# 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>.

1. Note that you have to first get the data (e.g., via elton pull EMTuckerLabUMMZ/ummzi) before being able to generate reviews (e.g., elton review EMTuckerLabUMMZ/ummzi), extract interaction claims (e.g., elton interactions EMTuckerLabUMMZ/ummzi), or list taxonomic names (e.g., elton names EMTuckerLabUMMZ/ummzi) [↑](#footnote-ref-31)
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. [↑](#footnote-ref-40)
3. Up-to-date status of the GloBI Review Badge can be retrieved from the [GloBI Review Depot](https://depot.globalbioticinteractions.org/reviews/EMTuckerLabUMMZ/ummzi/review.svg) [↑](#footnote-ref-134)
4. Up-to-date status of the GloBI Index Badge can be retrieved from [GloBI’s API](https://api.globalbioticinteractions.org/interaction.svg?interactionType=ecologicallyRelatedTo&accordingTo=globi:EMTuckerLabUMMZ/ummzi&refutes=true&refutes=false) [↑](#footnote-ref-141)
5. At time of writing (2025-04-10) the version of the GloBI dataset index was available at <https://globalbioticinteractions.org/datasets> [↑](#footnote-ref-143)
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.” [↑](#footnote-ref-147)