# Versioned Archive and Review of Biotic Interactions and Taxon Names Found within JoseBSL/EuPPollNet hash://md5/346321bf93ff80852d25f24efaafb18d

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

2025-07-07

**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 JoseBSL/EuPPollNet, has fingerprint hash://md5/346321bf93ff80852d25f24efaafb18d, is 279MiB in size and contains 623,476 interaction with 1 unique type of association (e.g., pollinatedBy) between 1,543 primary taxon (e.g., Helianthus annuus) and 2,668 associated taxon (e.g., Apis mellifera). This report includes detailed summaries of interaction data, a taxonomic review from multiple catalogs, and an archived version of the dataset from which the reviews are derived.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:

> Lanuza et al. (2025), EuPPollNet: A European Database of Plant-Pollinator Networks. Global Ecol Biogeogr, 34: e70000. https://doi.org/10.1111/geb.70000 https://github.com/JoseBSL/EuPPollNet/archive/1f17adc594de4c7e 2025-07-05T07:57:23.887Z hash://md5/346321bf93ff80852d25f24efaafb18d

For additional metadata related to this dataset, please visit https://github.com/JoseBSL/EuPPollNet 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 | 0.15.13 |
| nomer | 0.5.17 |
| globinizer | 0.4.0 |
| mlr | 6.0.0 |
| jq | 1.6 |
| yq | 4.25.3 |
| pandoc | 3.1.6.1 |
| duckdb | 1.3.1 |

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

```
# get versioned copy of the dataset (size approx.  279MiB) under review
elton pull JoseBSL/EuPPollNet

# generate review notes
elton review JoseBSL/EuPPollNet\
 > review.tsv

# export indexed interaction records
elton interactions JoseBSL/EuPPollNet\
 > interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names JoseBSL/EuPPollNet\
 | nomer append col\
 > name-alignment.tsv
```

or visually, in a process diagram.

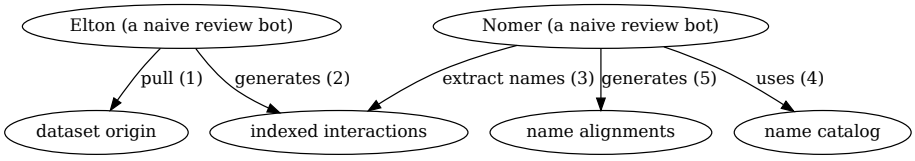


Figure 1: Review Process Overview

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

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 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.parquet | species interaction claims indexed from the dataset under review in Apache Parquet 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.parquet | taxonomic names found in the dataset under review in Apache Parquet format |

| filename | description |
|---|---|
| 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-col.parquet | 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 Apache Parquet 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 |

| filename | description |
|---|---|
| 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-discoverlife.parquet | 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 Apache Parquet 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-gbif.parquet | 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 Apache Parquet format |

| filename | description |
|---|---|
| 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-itis.parquet | 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 Apache Parquet 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 |

| filename | description |
|---|---|
| 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-mdd.parquet | 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 Apache Parquet 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-ncbi.parquet | 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 Apache Parquet 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-pbdb.parquet | 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 Apache Parquet 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 |

| filename | description |
|---|---|
| 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-tpt.parquet | 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 Apache Parquet 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-wfo.parquet | 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 Apache Parquet format |

| filename | description |
|---|---|
| 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-resolved-worms.parquet | 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 Apache Parquet 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) |

| filename | description |
|---|---|
| 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 organisms with specific taxa. The primary and associate organisms The kind of interaction is documented as an interaction type.

The dataset under review, named JoseBSL/EuPPollNet, has fingerprint

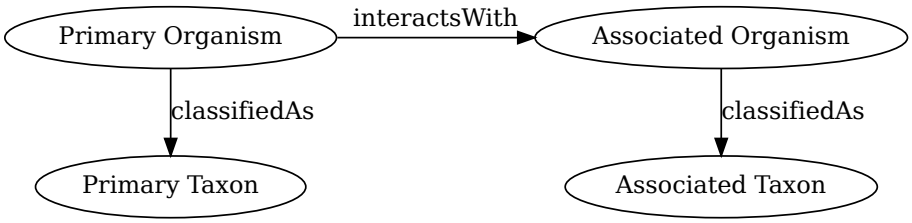


Figure 2: Biotic Interaction Data Model

hash://md5/346321bf93ff80852d25f24efaafb18d, is 279MiB in size and contains 623,476 interaction with 1 unique type of association (e.g., pollinatedBy) between 1,543 primary taxon (e.g., Helianthus annuus) and 2,668 associated taxon (e.g., Apis mellifera).

An exhaustive list of indexed interaction claims can be found in gzipped csv, tsv and parquet 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 |
|---|---|---|---|
| Urospermum picroides | pollinatedBy | Oedemera flavipes | doi:10.1007/s00442-007-0946-1 |
| Sonchus tenerrimus | pollinatedBy | Rhodanthidium sticticum | doi:10.1007/s00442-007-0946-1 |
| Sonchus tenerrimus | pollinatedBy | Oedemera lurida | doi:10.1007/s00442-007-0946-1 |
| Lavandula stoechas | pollinatedBy | Rhodanthidium sticticum | doi:10.1007/s00442-007-0946-1 |

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

| interactionTypeName | count |
|---|---|
| pollinatedBy | 623476 |

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

| sourceTaxonName | count |
|---|---|
| Helianthus annuus | 352151 |
| Trifolium repens | 9881 |
| Trifolium pratense | 7178 |
| Brassica oleracea | 6067 |
| Centaurea jacea | 5699 |
| Cistus crispus | 5065 |
| Lotus corniculatus | 4607 |
| Daucus carota | 4161 |
| Cirsium arvense | 4019 |
| Cistus monspeliensis | 3939 |
| Knautia arvensis | 3862 |
| Melilotus italicus | 3629 |
| Calluna vulgaris | 3362 |
| Centaurea scabiosa | 3212 |
| Ranunculus acris | 3137 |
| Achillea millefolium | 3037 |
| Lavandula stoechas | 2898 |
| Chaerophyllum hirsutum | 2782 |
| Thymbra capitata | 2702 |

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

| targetTaxonName | count |
|---|---|
| Apis mellifera | 406365 |
| Bombus terrestris | 21424 |
| Bombus pascuorum | 13825 |
| Bombus lapidarius | 10049 |
| Sphaerophoria scripta | 6570 |
| Episyrphus balteatus | 6504 |
| Bombus lucorum | 6084 |
| Eristalis tenax | 5823 |
| Halictus scabiosae | 4047 |
| Bombus soroeensis | 3847 |
| Bombus pratorum | 3558 |
| Bombus hortorum | 3494 |
| Bombus | 3411 |
| Anthophora | 3368 |
| Bombus wurflenii | 2792 |

| targetTaxonName | count |
|---|---|
| Syrphidae | 2772 |
| Anthophora plumipes | 2589 |
| Thricops beckeri | 2257 |
| Brassicogethes aeneus | 1876 |

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

| sourceTaxonName | interactionTypeName | targetTaxonName | count |
|---|---|---|---|
| Helianthus annuus | pollinatedBy | Apis mellifera | 342329 |
| Helianthus annuus | pollinatedBy | Bombus terrestris | 6658 |
| Trifolium repens | pollinatedBy | Apis mellifera | 5703 |
| Melilotus italicus | pollinatedBy | Apis mellifera | 3629 |
| Cistus crispus | pollinatedBy | Apis mellifera | 3366 |
| Brassica oleracea | pollinatedBy | Bombus terrestris | 2928 |
| Brassica oleracea | pollinatedBy | Anthophora | 2597 |
| Helianthus annuus | pollinatedBy | Halictus scabiosae | 2543 |
| Lavandula stoechas | pollinatedBy | Apis mellifera | 2257 |
| Thymbra capitata | pollinatedBy | Apis mellifera | 2022 |
| Chaerophyllum hirsutum | pollinatedBy | Thricops beckeri | 1981 |
| Trifolium pratense | pollinatedBy | Bombus pascuorum | 1873 |
| Cirsium arvense | pollinatedBy | Apis mellifera | 1655 |
| Tetrapanax papyrifer | pollinatedBy | Apis mellifera | 1614 |
| Brassica napus | pollinatedBy | Apis mellifera | 1547 |
| Centaurea jacea | pollinatedBy | Apis mellifera | 1529 |
| Rubus | pollinatedBy | Apis mellifera | 1450 |
| Melilotus albus | pollinatedBy | Apis mellifera | 1439 |
| Lotus corniculatus | pollinatedBy | Apis mellifera | 1413 |

**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.

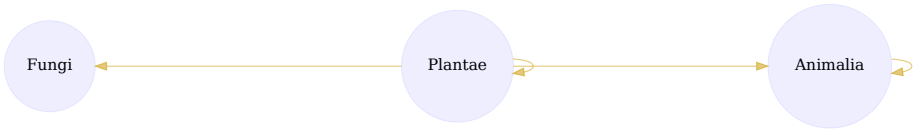


Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

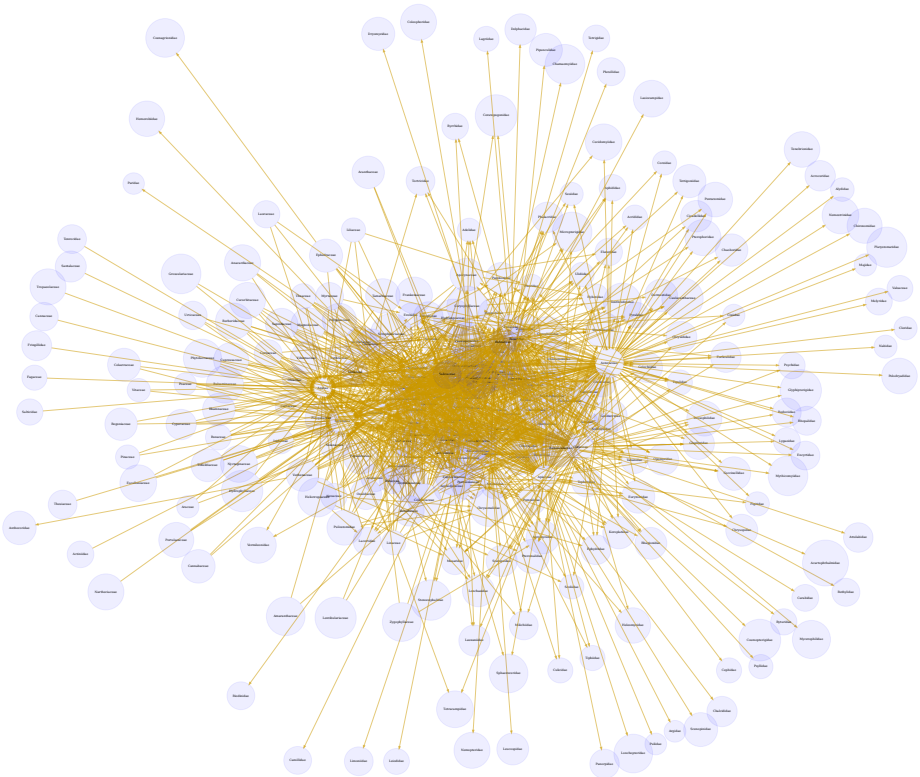

Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. download svg

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 |
|---|---|---|---|
| Abelia chinensis | HAS_ACCEPTED_NAME | col | Abelia chinensis |
| Abia fasciata | NONE | col | Abia fasciata |
| Abia nitens | NONE | col | Abia nitens |
| Abia sericea | NONE | col | Abia sericea |

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 | 367 |
| col | family | 104 |
| col | genus | 402 |
| col | order | 11 |
| col | species | 3314 |
| col | subfamily | 2 |
| col | subgenus | 7 |
| col | suborder | 1 |
| col | subspecies | 151 |
| col | superfamily | 1 |
| col | tribe | 2 |
| col | variety | 6 |
| discoverlife | NA | 3481 |
| discoverlife | species | 727 |
| gbif | NA | 25 |
| gbif | family | 113 |
| gbif | genus | 461 |
| gbif | order | 11 |
| gbif | species | 3596 |
| gbif | subspecies | 191 |

| resolvedCatalogName | resolvedRank | count |
|---|---|---|
| gbif | variety | 10 |
| itis | NA | 1829 |
| itis | family | 103 |
| itis | genus | 348 |
| itis | order | 11 |
| itis | species | 1892 |
| itis | subfamily | 5 |
| itis | suborder | 6 |
| itis | subspecies | 10 |
| itis | superfamily | 1 |
| itis | superorder | 1 |
| itis | variety | 6 |
| mdd | NA | 4208 |
| ncbi | NA | 598 |
| ncbi | cohort | 1 |
| ncbi | family | 106 |
| ncbi | genus | 420 |
| ncbi | order | 11 |
| ncbi | section | 1 |
| ncbi | species | 3050 |
| ncbi | subfamily | 8 |
| ncbi | subgenus | 19 |
| ncbi | suborder | 5 |
| ncbi | subspecies | 5 |
| ncbi | superfamily | 1 |
| ncbi | varietas | 1 |
| pbdb | NA | 3845 |
| pbdb | family | 109 |
| pbdb | genus | 200 |
| pbdb | infraclass | 1 |
| pbdb | infraorder | 1 |
| pbdb | order | 12 |
| pbdb | species | 26 |
| pbdb | subfamily | 7 |
| pbdb | suborder | 7 |
| pbdb | superfamily | 1 |
| pbdb | unranked clade | 1 |
| tpt | NA | 4204 |
| tpt | genus | 2 |
| tpt | species | 2 |
| wfo | NA | 2664 |
| wfo | family | 1 |
| wfo | genus | 147 |
| wfo | species | 1390 |

| resolvedCatalogName | resolvedRank | count |
|---|---|---|
| wfo | subspecies | 52 |
| wfo | variety | 2 |
| worms | NA | 3486 |
| worms | family | 83 |
| worms | genus | 185 |
| worms | order | 10 |
| worms | species | 436 |
| worms | suborder | 5 |
| worms | subspecies | 12 |

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 | 3811 |
| col | NONE | 367 |
| col | SYNONYM_OF | 938 |
| discoverlife | NONE | 3483 |
| discoverlife | HAS_ACCEPTED_NAME | 697 |
| discoverlife | SYNONYM_OF | 183 |
| discoverlife | HOMONYM_OF | 47 |
| gbif | HAS_ACCEPTED_NAME | 4835 |
| gbif | SYNONYM_OF | 1793 |
| gbif | NONE | 25 |
| itis | NONE | 1829 |
| itis | HAS_ACCEPTED_NAME | 2306 |
| itis | SYNONYM_OF | 163 |
| mdd | NONE | 4211 |
| ncbi | SAME_AS | 3566 |
| ncbi | NONE | 598 |
| ncbi | SYNONYM_OF | 92 |
| pbdb | NONE | 3846 |
| pbdb | HAS_ACCEPTED_NAME | 372 |
| pbdb | SYNONYM_OF | 9 |
| tpt | NONE | 4207 |
| tpt | HAS_ACCEPTED_NAME | 3 |
| tpt | SYNONYM_OF | 1 |

| resolvedCatalogName | relationName | count |
|---|---|---|
| wfo | SYNONYM_OF | 318 |
| wfo | NONE | 2664 |
| wfo | HAS_ACCEPTED_NAME | 1548 |
| wfo | HAS_UNCHECKED_NAME | 273 |
| worms | NONE | 3486 |
| worms | HAS_ACCEPTED_NAME | 850 |
| worms | SYNONYM_OF | 144 |

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 |
|---|---|---|
| 2025-07-07T16:27:52Z | note | source taxon name missing |
| 2025-07-07T16:27:52Z | note | source taxon name missing |
| 2025-07-07T16:27:52Z | note | source taxon name missing |
| 2025-07-07T16:27:52Z | note | source taxon name missing |

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

Table 13: Most frequently occurring review notes, if any.

| reviewComment | count |
|---|---|
| found malformed doi [NA] | 412951 |
| found malformed doi [Unpublished] | 28058 |
| found malformed doi [https://doi.org/10.1111/oik.08902; https://doi.org/10.1002/ecy.3712] | 14526 |
| found malformed doi [https://doi.org/10.1111/oik.07259; 10.1016/j.actao.2020.103551] | 5511 |

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

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

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.

# Acknowledgements

We thank the many humans that created us and those who created and maintained the data, software and other intellectual resources that were

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

[5] At time of writing (2025-07-07) 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, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."

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/ B60c0d25a16ae77b24305782017b1a270b79b5d1746f832650F2027ba536e276 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.