

A Review of Biotic Interactions and Taxon Names Found in Big-Bee-Network/select-bee-interactions.sh

by Nomer and Elton, two naive review bots
review@globalbioticinteractions.org
<https://globalbioticinteractions.org/contribute>
<https://github.com/Big-Bee-Network/select-bee-interactions.sh/issues>

2024-06-07

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 process of such an openly accessible digital interactions dataset of known origin, and discuss its outcome. The dataset under review, named Big-Bee-Network/select-bee-interactions.sh, is 388MiB in size and contains 270159 interactions with 22 unique types of associations (e.g., visitsFlowersOf) between 6695 primary taxa (e.g., Arthropoda) and 12333 associated taxa (e.g., Insecta). The report includes detailed summaries of interactions data as well as a taxonomic review from multiple catalogs.

Contents

Introduction	2
Data Review	2
Methods	2
Results	3
Biotic Interactions	3
Interaction Networks	8
Taxonomic Alignment	8
Additional Reviews	15
GloBI Review Badge	17
GloBI Index Badge	17
Discussion	17

Acknowledgements	18
Author contributions	18
References	18

Introduction

Data Review

Data review can be a time consuming process, especially when done manually. This review report aims to help facilitate data review of species interaction claims made in datasets registered with Global Biotic Interactions (Poelen, Simons, and Mungall 2014). The review includes summary statistics of, and observations about, the dataset under review:

Seltmann KC, Poelen JH (2024) Likely Apoidea (bees and wasps) Interactions Extracted from Global Biotic Interactions Verbatim Data Product using Nomer's DiscoverLife Support.
file:///home/runner/work/select-bee-interactions.sh/select-bee-interactions.sh/./

For additional metadata related to this dataset, please visit <https://github.com/Big-Bee-Network/select-bee-interactions.sh> 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, Elton, Nomer combined with third-party tools like grep, mlr, tail and head.

Table 1: Tools used in this review process

tool name	version
elton	0.13.4
nomer	0.5.9
mlr	6.0.0
pandoc	3.1.6.1

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

¹Note that you have to first get the data (e.g., via `elton pull Big-Bee-Network/select-bee-interactions.sh`) before being able to generate reviews (e.g., `elton review Big-Bee-Network/select-bee-interactions.sh`), extract interaction claims (e.g., `elton interactions Big-Bee-Network/select-bee-interactions.sh`), or list taxonomic names (e.g., `elton names Big-Bee-Network/select-bee-interactions.sh`)

```
# get versioned copy of the dataset (size approx. 388MiB) under review
elton pull Big-Bee-Network/select-bee-interactions.sh

# generate review notes
elton review Big-Bee-Network/select-bee-interactions.sh\
> review.tsv

# export indexed interaction records
elton interactions Big-Bee-Network/select-bee-interactions.sh\
> interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names Big-Bee-Network/select-bee-interactions.sh\
| nomer append col\
> name-alignment.tsv
```

or visually, in a process diagram.

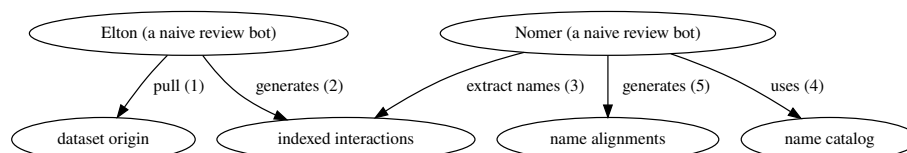


Figure 1: Review Process Overview

You can find a recent copy of the full review script at [check-data.sh](#).

Results

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

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 `Big-Bee-Network/select-bee-interactions.sh`, is 388MiB in size and contains 270159 interactions with 22 unique types of

²Disclaimer: 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.

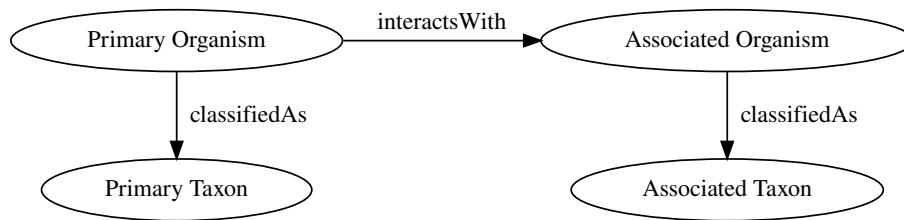


Figure 2: Biotic Interaction Data Model

associations (e.g., visitsFlowersOf) between 6695 primary taxa (e.g., Arthropoda) and 12333 associated taxa (e.g., Insecta).

An exhaustive list of indexed interaction claims can be found in csv and tsv archives. To facilitate discovery, the first 500 claims available on the html page at indexed-interactions.html are shown below.

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

Table 2: Sample of Indexed Interaction Claims

sourceTaxonName	interactionTypeNam	targetTaxonName	referenceCitation
Clausicella neomexicana	hemiparasiteOf	Bombus fervidus	Arnaud, Paul Henri. A Host-parasite Catalog of North American Tachinidae (Diptera). Washington, D.C.: U.S. Dept. of Agriculture, Science and Education Administration, 1978.

sourceTaxonName	interactionTypeNam	targetTaxonName	referenceCitation
Clausicella neomexicana	hemiparasiteOf	Bombus fervidus	Arnaud, Paul Henri. A Host-parasite Catalog of North American Tachinidae (Diptera). Washington, D.C.: U.S. Dept. of Agriculture, Science and Education Administration, 1978.
Tachina	hemiparasiteOf	Bombus	Arnaud, Paul Henri. A Host-parasite Catalog of North American Tachinidae (Diptera). Washington, D.C.: U.S. Dept. of Agriculture, Science and Education Administration, 1978.
Tachina	hemiparasiteOf	Bombus	Arnaud, Paul Henri. A Host-parasite Catalog of North American Tachinidae (Diptera). Washington, D.C.: U.S. Dept. of Agriculture, Science and Education Administration, 1978.

Table 3: Most Frequently Mentioned Interaction Types (up to 20 most frequent)

interactionTypeName	count
visitsFlowersOf	110701
hasHost	74676
eats	54090
interactsWith	27920
visits	10535
preysOn	4662
pollinates	4038
adjacentTo	955
endoparasiteOf	543
kleptoparasiteOf	332
parasiteOf	240
pathogenOf	230
coOccursWith	208
createsHabitatFor	93
hemiparasiteOf	68
parasitoidOf	55
hasVector	15
commensalistOf	14
ectoparasiteOf	14

Table 4: Most Frequently Mentioned Primary Taxa (up to 20 most frequent)

sourceTaxonName	count
Arthropoda	26238
Apis mellifera	24463
Andrena	19080
Acari	12880
Bombus impatiens	10747
Bombus	10160
Megachile	5670
Bombus griseocollis	4945
Bombus melanopygus	4043
Bombus pensylvanicus	3990
Bombus vosnesenskii	3378
Bombus flavifrons	3170
Bombus terrestris	2961
Pyrobombus	2800
Bombus bimaculatus	2788

sourceTaxonName	count
Bombus perplexus	2602
Hylaeus	2394
Halictus ligatus	2340
Halictus tripartitus	2256

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

targetTaxonName	count
Insecta	8230
Solidago	3805
Animalia	3769
Rubus	3289
Pycnanthemum muticum	3144
Apis mellifera	2880
Megachile	2595
Trifolium repens	2496
Monarda fistulosa	2439
Hymenoptera	2416
Arthropoda	2153
Heliomeris multiflora	2116
Symphotrichum	2094
Taraxacum officinale	1947
Centaurea stoebe	1919
Heterotheca villosa	1913
Cirsium	1759
Myodes rutilus	1671
Andrena	1620

Table 6: Most Frequent Interactions between Primary and Associate Taxa (up to 20 most frequent)

sourceTaxonName	interactionTypeName	targetTaxonName	count
Andrena	hasHost	Rubus	1740
Arthropoda	hasHost	Myodes rutilus	1671
Acari	hasHost	Megachile	1320
Bombus	hasHost	Centaurea stoebe	1130
Arthropoda	hasHost	Peromyscus keeni	1082
Arthropoda	hasHost	Microtus oeconomus	863
Vespa velutina	eats	Apis mellifera	836

sourceTaxonName	interactionTypeName	targetTaxonName	count
Bombus bifarius	visitsFlowersOf	Helioomeris multiflora	794
Bombus	hasHost	Solidago	780
Bombus flavifrons	visitsFlowersOf	Helioomeris multiflora	720
Bombus bifarius	visitsFlowersOf	Heterotheca villosa	626
Arthropoda	hasHost	Sorex cinereus	542
Andrena	visitsFlowersOf	Sonchus tenerrinus	540
Andrena	hasHost	Trifolium	540
Andrena	hasHost	Geranium maculatum	540
Bombus flavifrons	visitsFlowersOf	Heterotheca villosa	532
Apis mellifera	eats	Salvia rosmarinus	513
Acari	hasHost	Bombus	510
Halictus	hasHost	Achillea millefolium	486

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](#). A tab-separated file can be found at [indexed-interactions.tsv](#)

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

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

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, and pbdb). These alignments can help review name usage or aid in selecting of a suitable taxonomic name resource.

Table 7: Sample of Name Alignments

providedName	relationName	resolvedCatalogName	resolvedName
Anaphalis	HAS_ACCEPTED_BY	COL	Anaphalis
margaritacea			margaritacea
Agastache	HAS_ACCEPTED_BY	COL	Agastache
Aka	HAS_ACCEPTED_BY	COL	Aka

providedName	relationName	resolvedCatalogName	resolvedName
Aka	SYNONYM_OF	col	Siphonodictyon

Table 8: 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	2248
col	class	7
col	family	106
col	form	1
col	genus	1410
col	gigaclass	1
col	kingdom	2
col	order	28
col	phylum	3
col	section	1
col	species	12887
col	subfamily	33
col	subgenus	40
col	suborder	2
col	subspecies	675
col	subterclass	2
col	superfamily	2
col	superorder	1
col	tribe	14
col	variety	72
discoverlife	NA	14155
discoverlife	class	1
discoverlife	family	6
discoverlife	genus	108
discoverlife	kingdom	1
discoverlife	order	1
discoverlife	phylum	1
discoverlife	species	2885
discoverlife	subfamily	7
discoverlife	subgenus	147
discoverlife	subspecies	57
discoverlife	tribe	13
gbif	NA	1589
gbif	class	7

resolvedCatalogName	resolvedRank	count
gbif	family	109
gbif	form	9
gbif	genus	1519
gbif	kingdom	2
gbif	order	28
gbif	phylum	3
gbif	species	13295
gbif	subspecies	851
gbif	variety	212
itis	NA	3970
itis	class	7
itis	division	2
itis	family	108
itis	genus	1292
itis	infrakingdom	1
itis	kingdom	3
itis	order	27
itis	phylum	2
itis	species	11311
itis	subclass	1
itis	subfamily	10
itis	subgenus	1
itis	suborder	3
itis	subspecies	479
itis	superclass	1
itis	superfamily	2
itis	superorder	2
itis	tribe	13
itis	variety	65
mdd	NA	17286
ncbi	NA	4221
ncbi	clade	5
ncbi	class	7
ncbi	family	107
ncbi	genus	1373
ncbi	infraorder	1
ncbi	kingdom	1
ncbi	order	28
ncbi	phylum	2
ncbi	section	2
ncbi	species	11203
ncbi	species group	1
ncbi	subclass	1
ncbi	subfamily	36

resolvedCatalogName	resolvedRank	count
ncbi	subgenus	60
ncbi	suborder	1
ncbi	subspecies	176
ncbi	subtribe	10
ncbi	superclass	1
ncbi	superfamily	2
ncbi	superorder	2
ncbi	tribe	54
ncbi	varietas	16
pbdb	NA	14646
pbdb	class	10
pbdb	family	108
pbdb	genus	621
pbdb	kingdom	4
pbdb	order	30
pbdb	phylum	4
pbdb	species	1826
pbdb	subfamily	15
pbdb	suborder	4
pbdb	subspecies	3
pbdb	subtribe	1
pbdb	superclass	1
pbdb	superfamily	2
pbdb	superorder	1
pbdb	superphylum	1
pbdb	tribe	11
pbdb	unranked clade	10
tpt	NA	13496
tpt	family	1
tpt	genus	270
tpt	species	3519
wfo	NA	10639
wfo	class	1
wfo	family	77
wfo	form	1
wfo	genus	922
wfo	order	19
wfo	section	2
wfo	species	5526
wfo	subfamily	8
wfo	subsection	1
wfo	subspecies	101
wfo	subtribe	12
wfo	tribe	16

resolvedCatalogName	resolvedRank	count
wfo	variety	56

Table 9: 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	19457
col	SYNONYM_OF	3886
col	NONE	2480
discoverlife	NONE	17802
discoverlife	HAS_ACCEPTED_NAME	5144
discoverlife	SYNONYM_OF	1865
discoverlife	HOMONYM_OF	837
gbif	HAS_ACCEPTED_NAME	23266
gbif	SYNONYM_OF	6481
gbif	NONE	1794
itis	HAS_ACCEPTED_NAME	17631
itis	NONE	4333
itis	SYNONYM_OF	1577
mdd	NONE	21755
mdd	HAS_ACCEPTED_NAME	1357
ncbi	SAME_AS	17857
ncbi	NONE	4666
ncbi	COMMON_NAME_OF	4
ncbi	SYNONYM_OF	828
pbdb	NONE	19281
pbdb	HAS_ACCEPTED_NAME	3747
pbdb	SYNONYM_OF	178
tpt	NONE	18487
tpt	HAS_ACCEPTED_NAME	4621
tpt	SYNONYM_OF	7
wfo	HAS_ACCEPTED_NAME	8587
wfo	HAS_UNCHECKED_NAME	1259
wfo	NONE	13909
wfo	SYNONYM_OF	1739

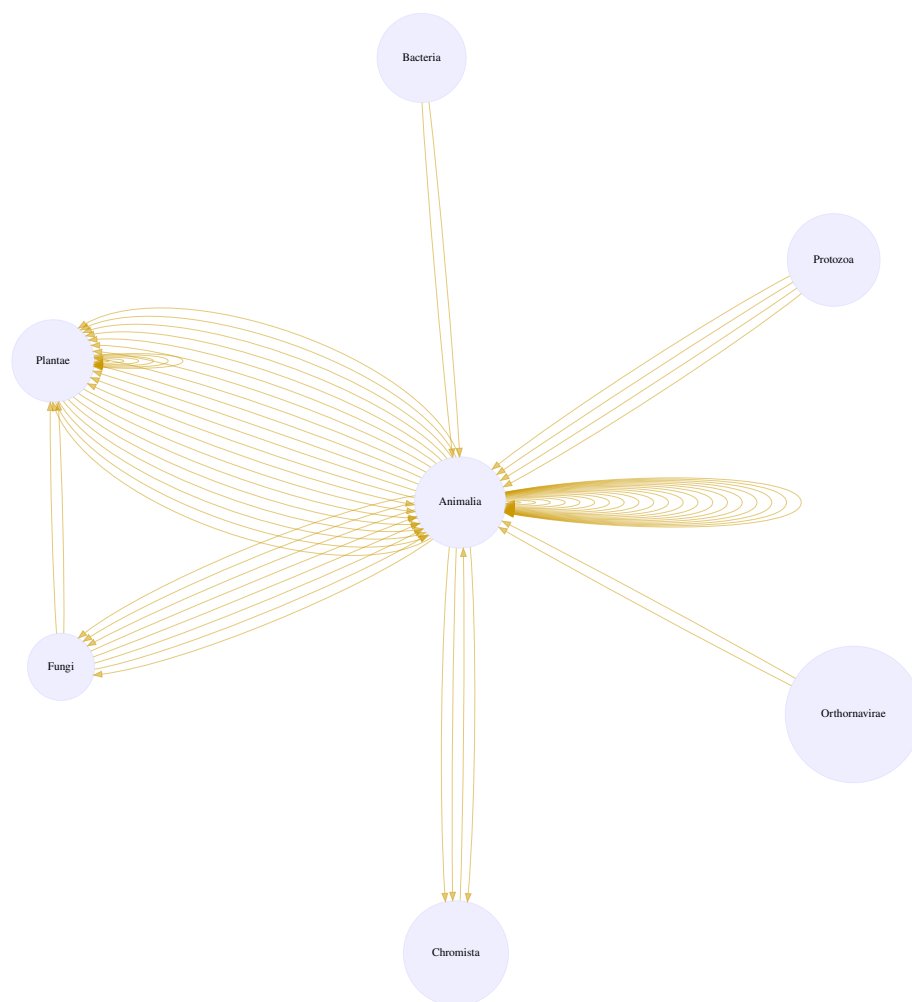


Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

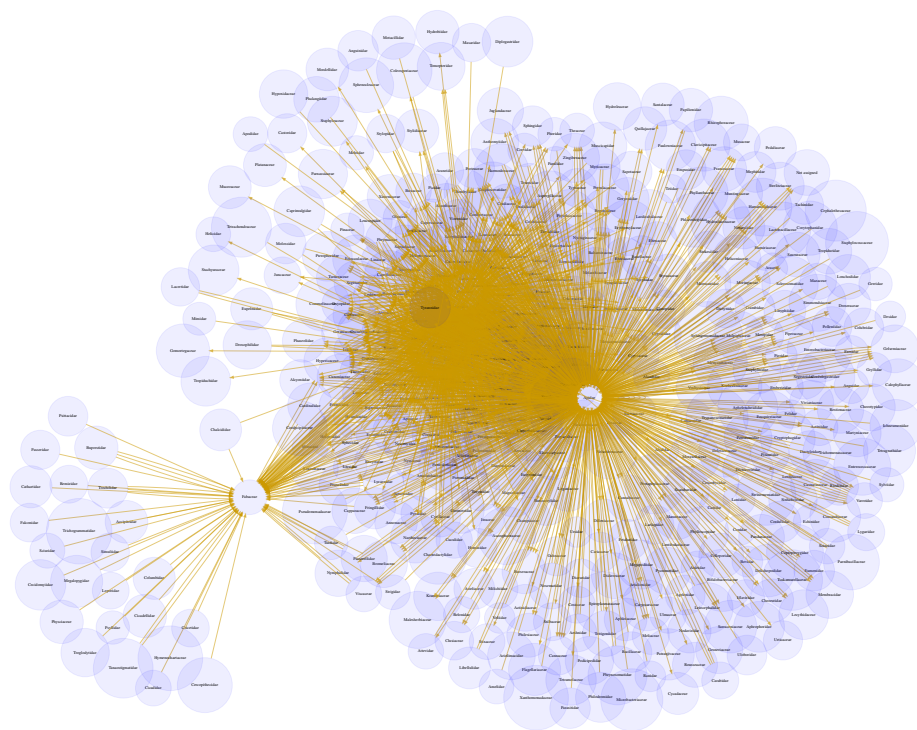


Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. [download svg](#)

Table 10: List of Available Name Alignment Reports

catalog name	alignment results
col	associated names alignments (first 500, full csv/tsv)
ncbi	associated names alignments (first 500, full csv/tsv)
discoverlife	associated names alignments (first 500, full csv/tsv)
gbif	associated names alignments (first 500, full csv/tsv)
itis	associated names alignments (first 500, full csv/tsv)
wfo	associated names alignments (first 500, full csv/tsv)
mdd	associated names alignments (first 500, full csv/tsv)
tpt	associated names alignments (first 500, full csv/tsv)
pbdb	associated names alignments (first 500, full csv/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 11: First few lines in the review notes.

reviewDate	reviewCommentType	reviewComment
2024-06-07T17:05:39Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geo not found]
2024-06-07T17:05:39Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geo not found]

reviewDate	reviewCommentType	reviewComment
2024-06-07T17:05:39Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]
2024-06-07T17:05:39Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]

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

Table 12: Most frequently occurring review notes, if any.

reviewComment	count
failed to lookup [GEONAMES:5539795] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5539795&username=globi&style=full] not found]	27711
failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]	752
failed to lookup [GEONAMES:3895114] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=3895114&username=globi&style=full] not found]	677
failed to lookup [GEONAMES:4896861] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=4896861&username=globi&style=full] not found]	528

For addition information on review notes, please have a look at the first 500 Review Notes or the download full 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 successfully 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 ^[At time of writing (2024-06-07) the version of the GloBI dataset index was available at <https://globalbioticinteractions.org/datasets> for badge examples.

Discussion

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-ness⁵ 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.

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

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

⁵According 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."

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 processing 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.

Author contributions

Nomer was responsible for name alignments. Elton carried out dataset extraction, and generated the review notes.

References

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