

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 51.0MiB in size and contains 24601 interactions with 18 unique types of associations (e.g., visitsFlowersOf) between 1754 primary taxa (e.g., *Andrena*) and 3689 associated taxa (e.g., *Heliomeris multiflora*). 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	14
GloBI Review Badge	15
GloBI Index Badge	16
Discussion	16

Acknowledgements	17
Author contributions	17
References	17

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. 51.0MiB) 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.

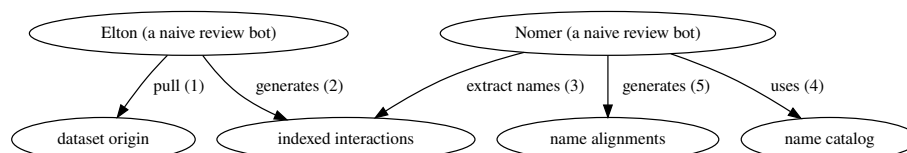


Figure 1: Review Process Overview

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 51.0MiB in size and contains 24601 interactions with 18 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.

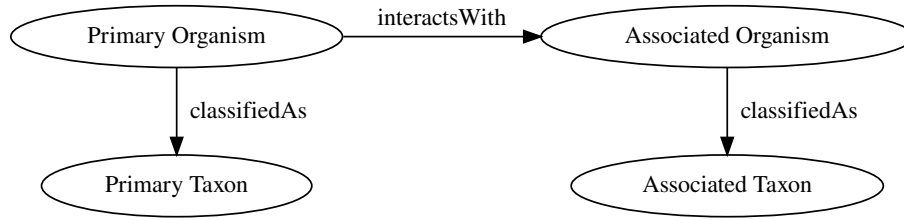


Figure 2: Biotic Interaction Data Model

associations (e.g., visitsFlowersOf) between 1754 primary taxa (e.g., *Andrena*) and 3689 associated taxa (e.g., *Helioeris multiflora*).

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	interactionTypeName	targetTaxonName	referenceCitation
Clausicella neomexicana	parasiteOf	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	parasiteOf	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.
Lespesia frenchii	parasiteOf	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.
Lespesia frenchii	parasiteOf	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	38345
hasHost	20656
interactsWith	2283
eats	2000
pollinates	1942
endoparasiteOf	543
visits	345
kleptoparasiteOf	166
pathogenOf	111
coOccursWith	104
parasiteOf	103
createsHabitatFor	93
adjacentTo	52
preysOn	19
parasitoidOf	19
commensalistOf	7
ectoparasiteOf	7
hasVector	6

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

sourceTaxonName	count
Andrena	8925
Acari	6284
Bombus	4045
Bombus flavifrons	1846
Bombus bifarius	1828
Apis mellifera	1678
Bombus lapidarius	1668
Megachile	1575
Bombus terrestris	1540
Apis	951
Vespa velutina	872
Ceratina	836
Halictus	828
Bombus pascuorum	640
Bombus hortorum	622
Bombus sylvicola	590

sourceTaxonName	count
Lasioglossum	548
Bombus appositus	512
Andrena cressonii	501

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

targetTaxonName	count
Heliomeris multiflora	2114
Heterotheca villosa	1912
Apis mellifera	1702
Rubus	1189
Megachile	840
Centaurea scabiosa	818
Centaurea stoebe	800
Solidago	658
Senecio integerrimus	634
Symphotrichum	629
Taraxacum officinale	594
Sonchus tenerrinus	579
Apis cerana	576
Andrena	570
Trifolium repens	536
Achillea millefolium	530
Delphinium nuttallianum	466
Cichorium intybus	453
Calluna vulgaris	453

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

sourceTaxonName	interactionTypeName	targetTaxonName	count
Andrena	hasHost	Rubus	870
Vespa velutina	eats	Apis mellifera	823
Bombus bifarius	visitsFlowersOf	Heliomeris multiflora	794
Bombus flavifrons	visitsFlowersOf	Heliomeris multiflora	720
Acari	hasHost	Megachile	660
Bombus bifarius	visitsFlowersOf	Heterotheca villosa	626
Bombus	hasHost	Centaurea stoebe	565
Andrena	visitsFlowersOf	Sonchus tenerrinus	540

sourceTaxonName	interactionTypeName	targetTaxonName	count
Bombus flavifrons	visitsFlowersOf	Heterotheca villosa	532
Bombus	hasHost	Solidago	390
Bombus lapidarius	visitsFlowersOf	Centaurea scabiosa	286
Bombus terrestris	visitsFlowersOf	Erica tetralix	271
Andrena	hasHost	Geranium maculatum	270
Andrena	hasHost	Trifolium	270
Acari	hasHost	Bombus	255
Halictus	hasHost	Achillea millefolium	243
Bombus	hasHost	Symphotrichum	240
Bombus terrestris	visitsFlowersOf	Erica cinerea	229
Andrena	visitsFlowersOf	Cistus monspeliensis	225

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
Milkweed	NONE	col	Milkweed
Squash	NONE	col	Squash
Thistle	NONE	col	Thistle
Acacia	HAS_ACCEPTED_NAME	col	Acacia

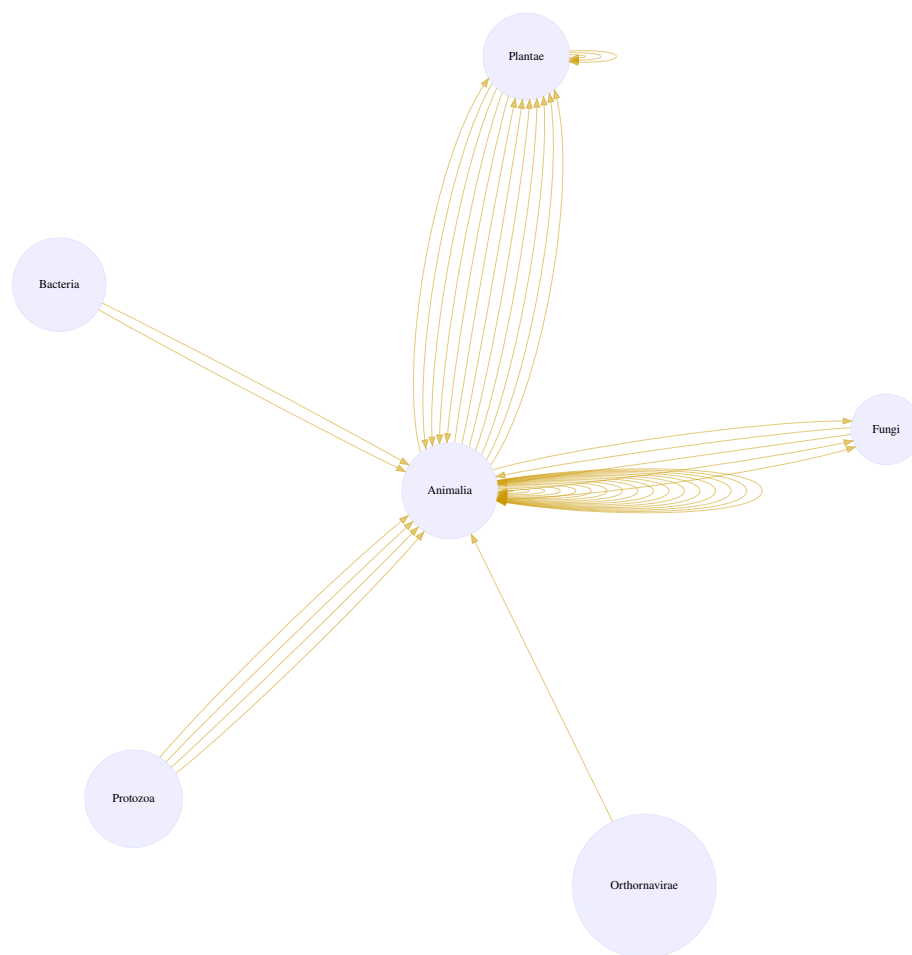


Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

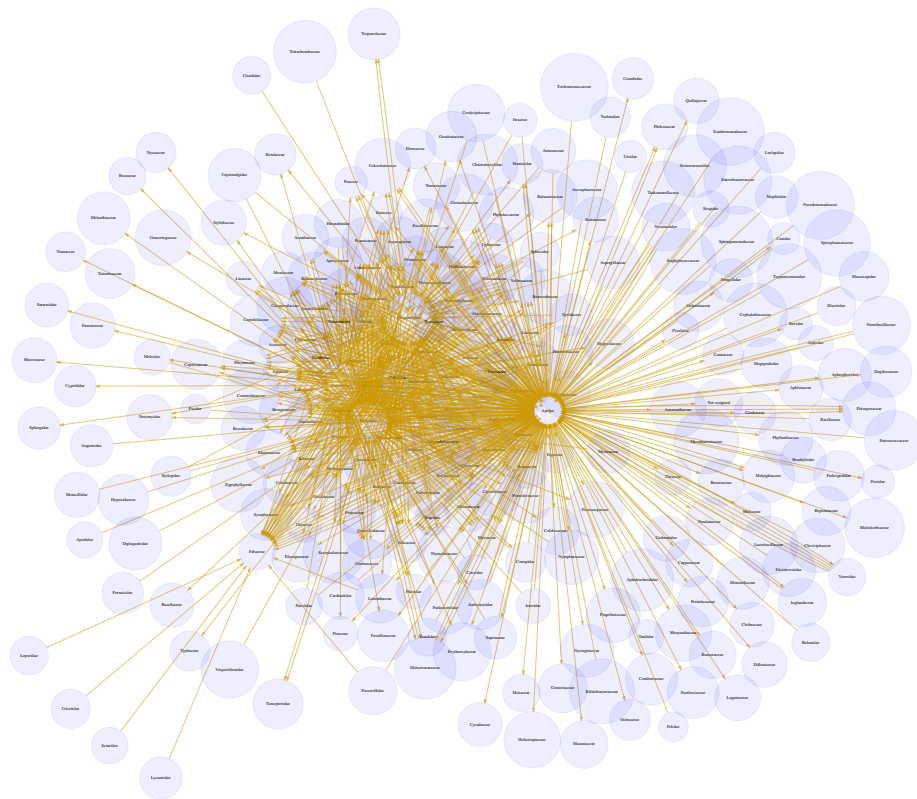


Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. [download svg](#)

Table 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	555
col	class	1
col	family	19
col	genus	355
col	gigaclass	1
col	phylum	1
col	section	1
col	species	4016
col	subfamily	3
col	subgenus	1
col	subspecies	158
col	tribe	3
col	variety	32
discoverlife	NA	2865
discoverlife	family	6
discoverlife	genus	75
discoverlife	phylum	1
discoverlife	species	2081
discoverlife	subfamily	2
discoverlife	subgenus	57
discoverlife	subspecies	33
discoverlife	tribe	3
gbif	NA	445
gbif	class	1
gbif	family	21
gbif	genus	379
gbif	phylum	1
gbif	species	4089
gbif	subspecies	195
gbif	variety	51
itis	NA	1067
itis	class	1
itis	family	21
itis	genus	338
itis	kingdom	1
itis	phylum	1
itis	species	3520
itis	subclass	1
itis	subfamily	2

resolvedCatalogName	resolvedRank	count
itis	subspecies	61
itis	superclass	1
itis	tribe	3
itis	variety	41
mdd	NA	5053
ncbi	NA	1530
ncbi	class	1
ncbi	family	20
ncbi	genus	342
ncbi	phylum	1
ncbi	section	1
ncbi	species	3110
ncbi	subclass	1
ncbi	subfamily	2
ncbi	subgenus	20
ncbi	subspecies	22
ncbi	superclass	1
ncbi	tribe	3
ncbi	varietas	10
pbdb	NA	4752
pbdb	class	3
pbdb	family	21
pbdb	genus	144
pbdb	kingdom	1
pbdb	phylum	1
pbdb	species	127
pbdb	subfamily	2
pbdb	suborder	1
pbdb	superclass	1
pbdb	tribe	3
pbdb	unranked clade	1
tpt	NA	4941
tpt	genus	5
tpt	species	107
wfo	NA	2801
wfo	family	11
wfo	genus	265
wfo	section	2
wfo	species	1941
wfo	subsection	1
wfo	subspecies	48
wfo	variety	23

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	NONE	601
col	HAS_ACCEPTED_NAME	5343
col	SYNONYM_OF	1550
discoverlife	NONE	3476
discoverlife	HAS_ACCEPTED_NAME	2911
discoverlife	SYNONYM_OF	1252
discoverlife	HOMONYM_OF	611
gbif	NONE	482
gbif	HAS_ACCEPTED_NAME	6636
gbif	SYNONYM_OF	2441
itis	NONE	1155
itis	HAS_ACCEPTED_NAME	4910
itis	SYNONYM_OF	672
mdd	NONE	6569
mdd	HAS_ACCEPTED_NAME	10
ncbi	NONE	1657
ncbi	SAME_AS	4729
ncbi	SYNONYM_OF	298
pbdb	NONE	6023
pbdb	HAS_ACCEPTED_NAME	548
pbdb	SYNONYM_OF	22
tpt	NONE	6466
tpt	HAS_ACCEPTED_NAME	111
tpt	SYNONYM_OF	3
wfo	NONE	3725
wfo	HAS_ACCEPTED_NAME	2473
wfo	SYNONYM_OF	800
wfo	HAS_UNCHECKED_NAME	562

Table 10: List of Available Name Alignment Reports

catalog name	alignment results
col	associated names alignments (first 500, full csv/tsv)

catalog name	alignment results
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)
tpst	associated names alignments (first 500, full csv/tsv)
pdbdb	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
2024-06-07T21:45:09Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geo not found]
2024-06-07T21:45:09Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geo not found]
2024-06-07T21:45:09Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geo not found]

reviewDate	reviewCommentType	reviewComment
2024-06-07T21:45:09Z	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: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]	511
failed to lookup [GEONAMES:7593737] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=7593737&username=globi&style=full] not found]	212

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.

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



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.

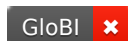


Figure 6: Picture of a GloBI Index Badge ⁴

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.

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 gener-

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."

ation 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>.