

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>

2024-06-08

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 1.18GiB in size and contains 769720 interactions with 21 unique types of associations (e.g., visitsFlowersOf) between 4813 primary taxa (e.g., *Apis mellifera*) and 11441 associated taxa (e.g., *Solidago*). 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	11
Additional Reviews	15
GloBI Review Badge	17
GloBI Index Badge	17
Discussion	18

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. 1.18GiB) 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.

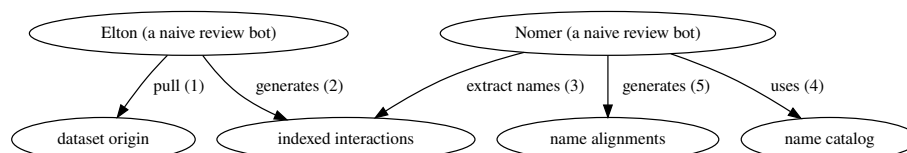


Figure 1: Review Process Overview

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 1.18GiB in size and contains 769720 interactions with 21 unique types of associ-

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.

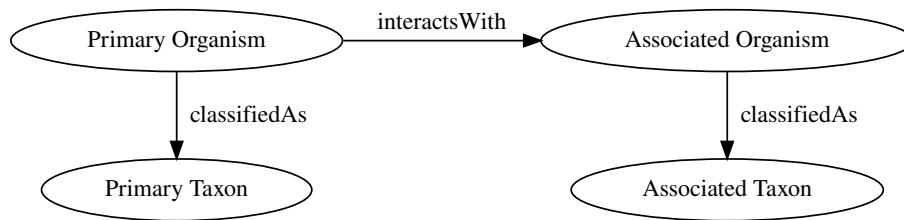


Figure 2: Biotic Interaction Data Model

ations (e.g., visitsFlowersOf) between 4813 primary taxa (e.g., *Apis mellifera*) and 11441 associated taxa (e.g., *Solidago*).

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	159240
pollinates	60835
eats	51858
hasHost	46000
interactsWith	29777
visits	14729
preysOn	3538
endoparasiteOf	543
kleptoparasiteOf	332
parasiteOf	254
pathogenOf	235
coOccursWith	208
adjacentTo	198
providesNutrientsFor	142
createsHabitatFor	93
parasitoidOf	60
ectoparasiteOf	14
hasVector	12
commensalistOf	9

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

sourceTaxonName	count
Apis mellifera	38921
Andrena	20760
Bombus impatiens	17733
Bombus	12680
Acari	12577
Bombus griseocollis	7272
Bombus vosnesenskii	5990
Megachile	5895
Bombus pensylvanicus	5762
Halictus	5229
Bombus melanopygus	5091
Bombus rufocinctus	4924
Bombus flavifrons	4806
Bombus bimaculatus	4495
Pyrobombus	4200

sourceTaxonName	count
Bombus perplexus	3982
Bombus bifarius	3920
Bombus terrestris	3903
Bombus pascuorum	3860

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

targetTaxonName	count
Solidago	6031
Trifolium repens	5077
Monarda fistulosa	3641
Rubus	3513
Symphyotrichum	3425
Trifolium pratense	3396
Pycnanthemum muticum	3264
Heliomeris multiflora	3192
Apis mellifera	3177
Megachile	2910
Heterotheca villosa	2874
Taraxacum officinale	2599
Taraxacum	2512
Salvia	2309
Nephrosperma vanhoutteanum	2301
Echinacea purpurea	2236
Cirsium	2019
Achillea millefolium	1999
Centaurea stoebe	1992

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

sourceTaxonName	interactionType	targetTaxonName	count
Andrena	hasHost	Rubus	1740
Lasioglossum mahense	pollinates	Nephrosperma vanhoutteanum	1441
Acari	hasHost	Megachile	1320
Bombus	hasHost	Centaurea stoebe	1130
Apis mellifera	pollinates	Trifolium repens	960

sourceTaxonName	interactionTypeNam	targetTaxonName	count
Bombus impatiens	visitsFlowersOf	Solidago	918
Vespa velutina	eats	Apis mellifera	838
Apis mellifera	pollinates	Nephrosperma vanhoutteanum	825
Bombus bifarius	visitsFlowersOf	Heliomeris multiflora	794
Bombus	hasHost	Solidago	780
Bombus flavifrons	visitsFlowersOf	Heliomeris multiflora	720
Apis mellifera	eats	Salvia rosmarinus	631
Bombus bifarius	visitsFlowersOf	Heterotheca villosa	626
Bombus impatiens	visitsFlowersOf	Symphyotrichum	585
Bombus vosnesenskii	visitsFlowersOf	Acmispon glaber	562
Andrena	visitsFlowersOf	Sonchus tenerrinus	540
Andrena	hasHost	Geranium maculatum	540
Andrena	hasHost	Trifolium	540
Bombus flavifrons	visitsFlowersOf	Heterotheca villosa	532

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](#).

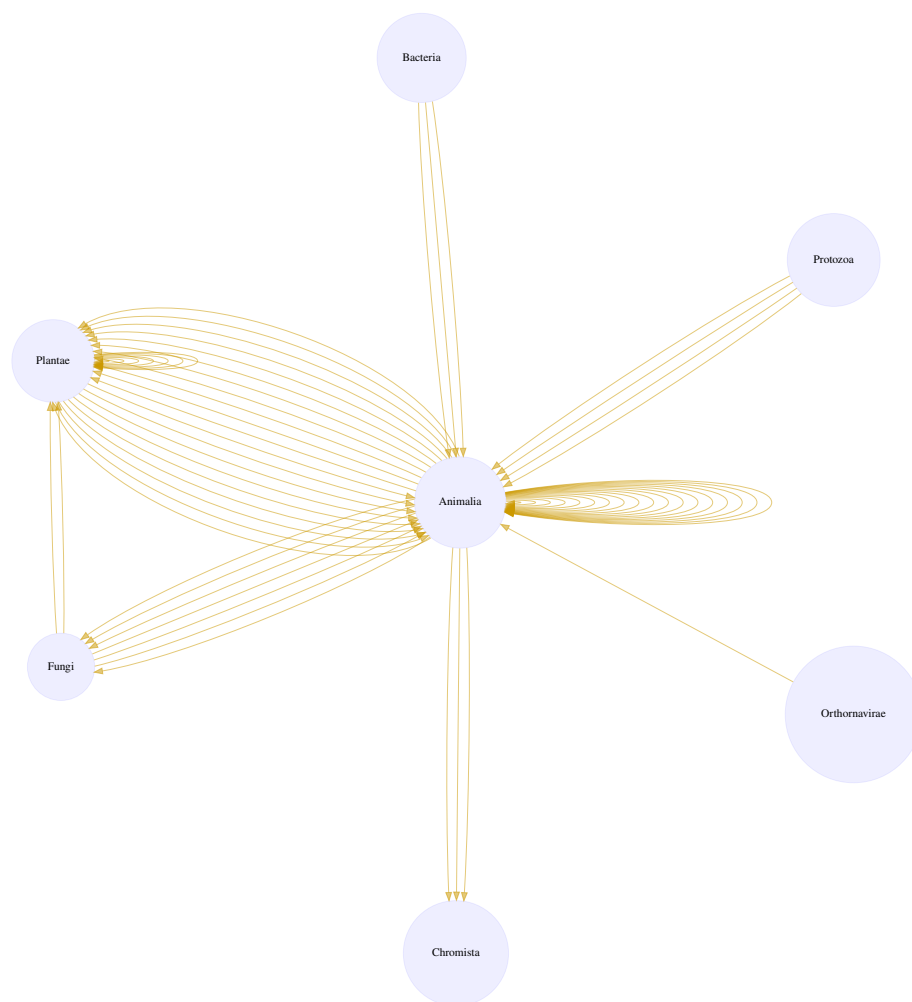


Figure 3: Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life download svg

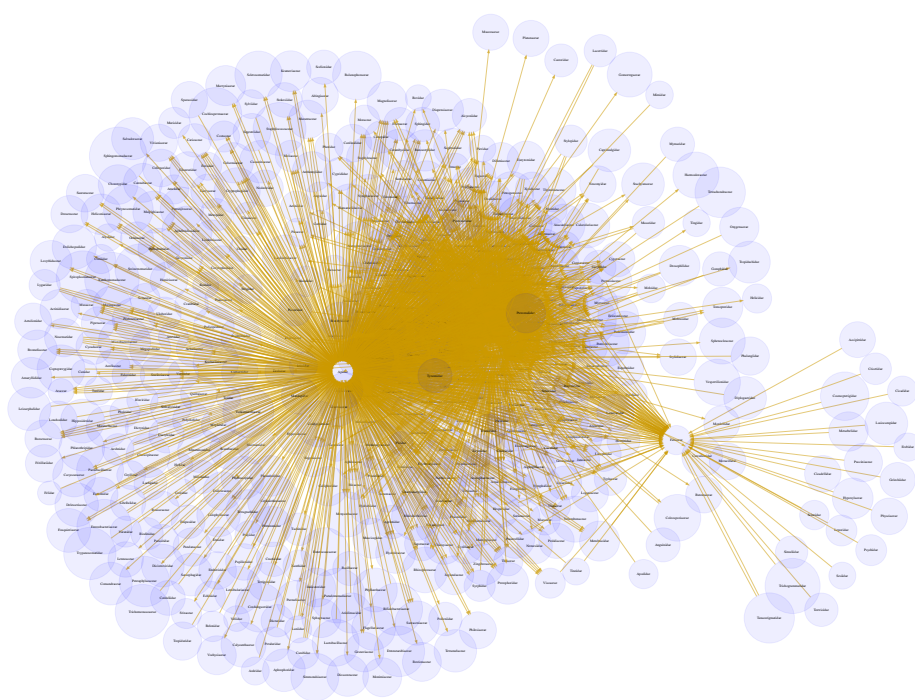


Figure 4: Interactions on the taxonomic family rank as interpreted by the Catalogue of Life. [download svg](#)

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
Agastache	HAS_ACCEPTED_NAME	col	Agastache
Aka	HAS_ACCEPTED_NAME	col	Aka
Aka	SYNONYM_OF	col	Siphonodictyon
Allium	HAS_ACCEPTED_NAME	col	Allium

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	1901
col	class	5
col	family	109
col	form	1
col	genus	1278
col	gigaclass	1
col	kingdom	1
col	order	29
col	phylum	4
col	section	1
col	species	10804
col	subfamily	38
col	subgenus	9
col	suborder	1
col	subspecies	388
col	subterclass	2
col	superfamily	2
col	superorder	1
col	tribe	15
col	variety	87
discoverlife	NA	10615
discoverlife	class	1
discoverlife	family	6

resolvedCatalogName	resolvedRank	count
discoverlife	genus	119
discoverlife	order	1
discoverlife	phylum	1
discoverlife	species	3525
discoverlife	subfamily	8
discoverlife	subgenus	168
discoverlife	subspecies	60
discoverlife	tribe	15
gbif	NA	1185
gbif	class	5
gbif	family	112
gbif	form	6
gbif	genus	1395
gbif	kingdom	1
gbif	order	29
gbif	phylum	4
gbif	species	11206
gbif	subspecies	616
gbif	variety	271
itis	NA	4258
itis	class	6
itis	division	3
itis	family	111
itis	genus	1167
itis	infrakingdom	1
itis	kingdom	2
itis	order	29
itis	phylum	1
itis	species	8598
itis	subclass	1
itis	subfamily	12
itis	subgenus	2
itis	suborder	2
itis	subspecies	139
itis	superclass	1
itis	superfamily	2
itis	superorder	2
itis	tribe	15
itis	variety	78
mdd	NA	14414
ncbi	NA	3995
ncbi	clade	6
ncbi	class	5
ncbi	family	109

resolvedCatalogName	resolvedRank	count
ncbi	genus	1237
ncbi	infraorder	1
ncbi	order	29
ncbi	phylum	2
ncbi	section	2
ncbi	species	8777
ncbi	species group	1
ncbi	subclass	1
ncbi	subfamily	41
ncbi	subgenus	61
ncbi	subspecies	61
ncbi	subtribe	12
ncbi	superclass	1
ncbi	superfamily	2
ncbi	superorder	2
ncbi	tribe	68
ncbi	varietas	21
pbdb	NA	13584
pbdb	class	8
pbdb	family	110
pbdb	genus	413
pbdb	kingdom	2
pbdb	order	32
pbdb	phylum	5
pbdb	species	224
pbdb	subfamily	16
pbdb	suborder	3
pbdb	subtribe	1
pbdb	superclass	1
pbdb	superfamily	2
pbdb	superorder	1
pbdb	superphylum	1
pbdb	tribe	13
pbdb	unranked clade	9
tpt	NA	14246
tpt	genus	8
tpt	species	160
wfo	NA	6194
wfo	class	1
wfo	family	89
wfo	form	1
wfo	genus	1056
wfo	order	22
wfo	section	2

resolvedCatalogName	resolvedRank	count
wfo	species	6925
wfo	subfamily	8
wfo	subsection	1
wfo	subspecies	129
wfo	subtribe	12
wfo	tribe	17
wfo	variety	69

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	17929
col	SYNONYM_OF	4372
col	NONE	2129
discoverlife	NONE	14561
discoverlife	HAS_ACCEPTED_NAME	6285
discoverlife	SYNONYM_OF	2319
discoverlife	HOMONYM_OF	935
gbif	HAS_ACCEPTED_NAME	22509
gbif	NONE	1376
gbif	SYNONYM_OF	7707
itis	HAS_ACCEPTED_NAME	15469
itis	NONE	4724
itis	SYNONYM_OF	1450
mdd	NONE	21100
mdd	HAS_ACCEPTED_NAME	15
ncbi	NONE	4483
ncbi	SAME_AS	16232
ncbi	COMMON_NAME_OF	3
ncbi	SYNONYM_OF	664
pbdb	NONE	19469
pbdb	HAS_ACCEPTED_NAME	1645
pbdb	SYNONYM_OF	70
tpt	NONE	20930
tpt	HAS_ACCEPTED_NAME	182
tpt	SYNONYM_OF	6

resolvedCatalogName	relationName	count
wfo	HAS_ACCEPTED_NAME	11219
wfo	NONE	9148
wfo	SYNONYM_OF	2249
wfo	HAS_UNCHECKED_NAME	1617

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
2024-06-08T15:18:35Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]
2024-06-08T15:18:35Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]
2024-06-08T15:18:35Z	note	failed to lookup [GEONAMES:5434527] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=5434527&username=globi&style=full] not found]
2024-06-08T15:18:35Z	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

reviewComment	count
failed to lookup [GEONAMES:4896861] because of: [resource [http://api.geonames.org/getJSON?formatted=true&geonameId=4896861&username=globi&style=full] not found]	511

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.

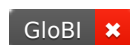


Figure 6: Picture of a GloBI Index Badge ⁴

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 (2024-06-08) the version of the GloBI dataset index was available at <https://globalbioticinteractions.org/datasets> for badge examples.

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

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 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

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.”

- 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>.