

# A Review of Biotic Interactions and Taxon Names Found in [globalbioticinteractions/carvalho2023](https://github.com/globalbioticinteractions/carvalho2023)

by Nomer and Elton, two naive review bots

[review@globalbioticinteractions.org](mailto:review@globalbioticinteractions.org)

<https://globalbioticinteractions.org/contribute>

<https://github.com/globalbioticinteractions/carvalho2023/issues>

2024-01-18

## 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 describe a review process of such an openly accessible digital interaction datasets of known origin, and discuss their outcome. The dataset under review (aka [globalbioticinteractions/carvalho2023](https://github.com/globalbioticinteractions/carvalho2023)) has size 327KiB and contains 542 interactions with 1 (e.g., `flowersVisitedBy`) unique types of associations between 63 primary taxa (e.g., *Scabiosa columbaria*) and 171 associated taxa (e.g., *Bombus pascuorum*). The report includes detailed summaries of interactions data as well as a taxonomic review from multiple perspectives.

## Table of Contents

Introduction .....	2
Data Review .....	2
Methods .....	2
Results .....	3
Biotic Interactions .....	4
Interaction Networks .....	7
Taxonomic Alignment .....	8
Additional Reviews .....	12
GloBI Review Badge .....	12
GloBI Index Badge .....	13
Discussion .....	13
Acknowledgements .....	14

## 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:

WorldFAIR pilot data from: VisitationData\_Luisa\_Carvalho.  
file:///home/runner/work/carvalho2023/carvalho2023/./

For additional metadata related to this dataset, please visit <https://github.com/globalbioticinteractions/carvalho2023> 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.

### *Tools used in this review process*

tool name	version
<a href="#">elton</a>	0.13.2
<a href="#">nomer</a>	0.5.6
<a href="#">mlr</a>	6.0.0
<a href="#">pandoc</a>	3.1.6.1

The review process can be described in the form of the script below <sup>1</sup>.

---

<sup>1</sup> Note that you have to first get the data (e.g., via `elton pull globalbioticinteractions/carvalho2023`) before being able to generate reviews (e.g., `elton review globalbioticinteractions/carvalho2023`), extract interaction claims (e.g., `elton interactions globalbioticinteractions/carvalho2023`), or list taxonomic names (e.g., `elton names globalbioticinteractions/carvalho2023`)

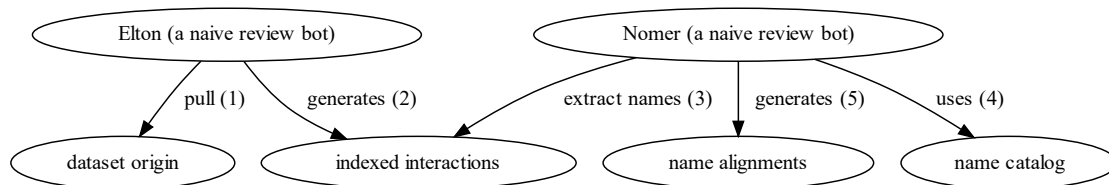
```
# get versioned copy of the dataset (size approx. 327KiB) under review
elton pull globalbioticinteractions/carvalho2023

# generate review notes
elton review globalbioticinteractions/carvalho2023\
> review.tsv

# export indexed interaction records
elton interactions globalbioticinteractions/carvalho2023\
> interactions.tsv

# export names and align them with the Catalogue of Life using Nomer
elton names globalbioticinteractions/carvalho2023\
| nomer append col\
> name-alignment.tsv
```

or visually, in a process diagram.



### Review Process Overview

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

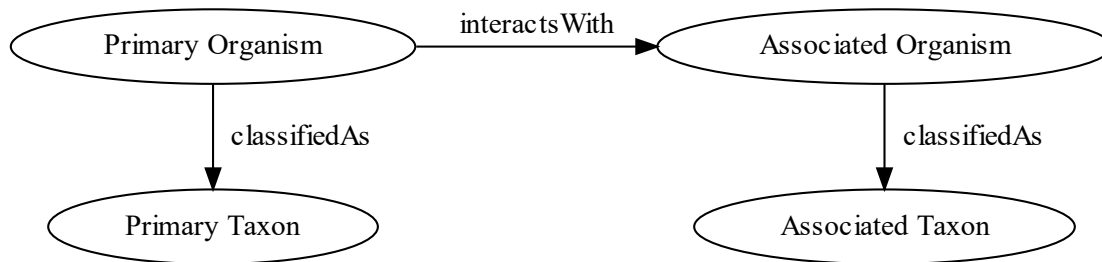
## Results

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

---

<sup>2</sup> 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.

## Biotic Interactions



### *Biotic Interaction Data Model*

In this review, biotic interactions (or biotic associations) are modeled as a primary (aka subject, source) organism interacting with an associate (aka object, target) organism. The dataset under review classified the primary/associate organisms with specific taxa. The primary and associate organisms The kind of interaction is documented as an interaction type.

The dataset under review (aka [globalbioticinteractions/carvalho2023](#)) has size 327KiB and contains 542 interactions with 1 (e.g., [flowersVisitedBy](#)) unique types of associations between 63 primary taxa (e.g., [Scabiosa columbaria](#)) and 171 associated taxa (e.g., [Bombus pascuorum](#)).

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.

### *Sample of Indexed Interaction Claims*

sourceTaxonName	interactionTypeName	targetTaxonName	referenceCitation
Pilosella officinarum	flowersVisitedBy	Andrena	<a href="https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546">https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546</a>
Mycelis muralis	flowersVisitedBy	Baccha	<a href="https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546">https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546</a>

sourceTaxonName	interactionTypeName	targetTaxonName	referenceCitation
			29546
Apiaceae	flowersVisitedBy	Bombus	<a href="https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546">https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546</a>
Origanum vulgare	flowersVisitedBy	Bombus	<a href="https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546">https://docs.google.com/spreadsheets/u/1/d/1cJ0qX9ppqHoSyqFykwYJefDFOzoutthBXjwKRY81T8/export?format=tsv&amp;gid=776329546</a>

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

interactionTypeName	count
flowersVisitedBy	685

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

sourceTaxonName	count
Scabiosa columbaria	44
Solidago virgaurea	39
Rubus fruticosus	36
Helianthemum nummularium	32
Geranium robertianum	32
Senecio	32
Crepis	30
Centranthus ruber	26
Clematis vitalba	24
Hieracium	23
Origanum vulgare	22
Achillea millefolium	22

sourceTaxonName	count
Pilosella officinarum	20
Torilis japonica	15
Cotoneaster franchetti	14
Hypericum perforatum	13
Ligustrum vulgare	13
Lactuca serriola	12
Eupatorium cannabinum	11

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

targetTaxonName	count
Bombus pascuorum	64
Episyrphus balteatus	25
Lasioglossum morio	22
Formica fusca	21
Meligethes solidus	20
Diptera	17
Syrphus ribesii	17
Heteroptera	16
Lasius alienus	15
Platycheirus albimanus	14
Hymenoptera	13
Halictus tumulorum	13
Anaspis pulicaria	12
Meligethes aeneus	10
Oedemera lurida	10
Paragus haemorrhous	10
Sphaerophoria scripta	10
Coleoptera	9
Thysanoptera	9

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

sourceTaxonName	interactionTypeName	targetTaxonName	count
Geranium robertianum	flowersVisitedBy	Bombus pascuorum	6

sourceTaxonName	interactionTypeName	targetTaxonName	count
Scabiosa columbaria	flowersVisitedBy	Bombus pascuorum	5
Solidago virgaurea	flowersVisitedBy	Bombus pascuorum	5
Teucrium scorodonia	flowersVisitedBy	Bombus pascuorum	5
Clinopodium ascendens	flowersVisitedBy	Bombus pascuorum	4
Lotus corniculatus	flowersVisitedBy	Bombus pascuorum	4
Origanum vulgare	flowersVisitedBy	Bombus pascuorum	4
Rubus fruticosus	flowersVisitedBy	Bombus pascuorum	4
Achillea millefolium	flowersVisitedBy	Heteroptera	3
Centranthus ruber	flowersVisitedBy	Homoptera	3
Solidago virgaurea	flowersVisitedBy	Sarcophaga	3
Rubus fruticosus	flowersVisitedBy	Anaspis pulicaria	3
Clematis vitalba	flowersVisitedBy	Episyrphus balteatus	3
Geranium robertianum	flowersVisitedBy	Episyrphus balteatus	3
Scabiosa columbaria	flowersVisitedBy	Episyrphus balteatus	3
Scabiosa columbaria	flowersVisitedBy	Eristalis tenax	3
Cotoneaster franchetti	flowersVisitedBy	Formica fusca	3
Clematis vitalba	flowersVisitedBy	Lasioglossum morio	3
Origanum vulgare	flowersVisitedBy	Lasioglossum morio	3

### 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. Note that both network graphs were first aligned taxonomically via the Catalogue of Life. Please refer to the original (or verbatim) taxonomic names for a more original view on the interaction data.



*Interactions on taxonomic kingdom rank as interpreted by the Catalogue of Life*  
[download svg](#)

As part of the review, all names are aligned against various name catalogs (e.g., `col ncbi discoverlife gbif itis globi mdd tpt pbdb`). These alignments may serve as a way to review name usage or aid in selecting of a suitable taxonomic name resource to use.



*Sample of Name Alignments*

providedName	relationName	resolvedCatalogName	resolvedName
Achillea millefolium	SYNONYM_OF	col	Achillea aspleniifolia
Achillea millefolium	HAS_ACCEPTED_NAME	col	Achillea millefolium
Achillea millefolium	SAME_AS	ncbi	Achillea millefolium
Achillea millefolium	NONE	discoverlife	Achillea millefolium

*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
tpt	NA	233
tpt	species	1
pbdb	NA	197
pbdb	genus	17
pbdb	family	9
pbdb	order	5
pbdb	species	3
pbdb	suborder	2
pbdb	subfamily	1
pbdb	infraorder	1
ncbi	species	156
ncbi	NA	36
ncbi	genus	22
ncbi	family	9
ncbi	subgenus	6
ncbi	order	5
ncbi	subspecies	2
ncbi	suborder	2
ncbi	subfamily	1
mdd	NA	234
itis	NA	103

resolvedCatalogName	resolvedRank	count
itis	species	94
itis	genus	19
itis	family	9
itis	order	5
itis	suborder	2
itis	subspecies	1
itis	subfamily	1
globi	species	161
globi	NA	159
globi	genus	50
globi	family	9
globi	subspecies	7
globi	order	5
globi	subgenus	4
globi	suborder	2
globi	subfamily	1
globi	infraorder	1
gbif	species	161
gbif	NA	32
gbif	genus	30
gbif	family	9
gbif	subspecies	7
gbif	order	5
discoverlife	NA	213
discoverlife	species	21
col	species	134
col	NA	61
col	genus	22
col	family	9
col	subspecies	6
col	order	5
col	subfamily	1

*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	SYNONYM_OF	42
col	HAS_ACCEPTED_NAME	165
col	NONE	61
discoverlife	NONE	213
discoverlife	HAS_ACCEPTED_NAME	21
discoverlife	HOMONYM_OF	2
discoverlife	SYNONYM_OF	5
gbif	HAS_ACCEPTED_NAME	230
gbif	SYNONYM_OF	72
gbif	NONE	32
globi	SAME_AS	2719
globi	NONE	21
itis	HAS_ACCEPTED_NAME	128
itis	NONE	103
itis	SYNONYM_OF	5
mdd	NONE	234
ncbi	SAME_AS	189
ncbi	SYNONYM_OF	14
ncbi	NONE	36
pbdb	NONE	197
pbdb	HAS_ACCEPTED_NAME	39
tpt	NONE	233
tpt	SYNONYM_OF	1

#### *List of Available Name Alignment Reports*

catalog name	alignment results
col	<a href="#">associated names alignments (first 500, full csv/tsv)</a>
ncbi	<a href="#">associated names alignments (first 500, full csv/tsv)</a>
discoverlife	<a href="#">associated names alignments (first 500, full csv/tsv)</a>
gbif	<a href="#">associated names alignments (first 500,</a>

catalog name	alignment results
	full <a href="#">csv</a> / <a href="#">tsv</a> )
itis	<a href="#">associated names alignments (first 500, full <a href="#">csv</a>/<a href="#">tsv</a>)</a>
globi	<a href="#">associated names alignments (first 500, full <a href="#">csv</a>/<a href="#">tsv</a>)</a>
mdd	<a href="#">associated names alignments (first 500, full <a href="#">csv</a>/<a href="#">tsv</a>)</a>
tpt	<a href="#">associated names alignments (first 500, full <a href="#">csv</a>/<a href="#">tsv</a>)</a>
pbdb	<a href="#">associated names alignments (first 500, full <a href="#">csv</a>/<a href="#">tsv</a>)</a>

## Additional Reviews

Elton, Nomer, and other tools may have difficulties interpreting existing species interaction datasets. Or, they may misbehave, or otherwise show unexpected behavior. As part of the review process, detailed review notes are kept that document possibly misbehaving, or confused, review bots. An sample of review notes associated with this review can be found below.

*First few lines in the review notes.*

reviewDate	reviewCommentType	reviewComment
2024-01-18T22:46:08Z	summary	file:///home/runner/work/carvalho2023/carvalho2023/./
2024-01-18T22:46:08Z	summary	685 interaction(s)
2024-01-18T22:46:08Z	summary	0 note(s)
2024-01-18T22:46:08Z	summary	685 info(s)

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

: Most frequently occurring review notes, if any.

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.



### *Sample of a GloBI Review Badge <sup>3</sup>*

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.



### *Sample of a GloBI Index Badge <sup>4</sup>*

If you'd like to keep track of reviews or index status of the dataset under review, please visit [GloBI's dataset index](#) <sup>5</sup> for badge examples.

## **Discussion**

This review is intended to provide a perspective on the dataset to aid understanding of species interaction claims discovered. However, this review should *not* be considered as fitness of use or other kind of quality assessment. Instead, the review may be used as an indication of the open-ness<sup>6</sup> and FAIRness (Wilkinson et al. 2016; Trekels et al. 2023) of the dataset: in order to perform this review, the data was likely openly available, **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. Currently, this Open-FAIR assessment is qualitative, and with measurement units specified, a more quantitative approach can be implemented.

---

<sup>3</sup> Up-to-date status of the GloBI Review Badge can be retrieved from the [GloBI Review Depot](#)

<sup>4</sup> Up-to-date status of the GloBI Index Badge can be retrieved from [GloBI's API](#)

<sup>5</sup> At time of writing (2024-01-18) the version of the GloBI dataset index was available at [\[https://globalbioticinteractions.org/datasets\]](https://globalbioticinteractions.org/datasets)(<https://globalbioticinteractions.org/datasets>)

<sup>6</sup> According to <http://opendefinition.org/>: "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."

## 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.

## 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>.