

A Review of Biotic Interactions and Taxon Names Found in [globalbioticinteractions/hivetracks2023](https://globalbioticinteractions.org/hivetracks2023)

by Nomer and Elton, two naive review bots

review@globalbioticinteractions.org

<https://globalbioticinteractions.org/contribute>

<https://github.com/globalbioticinteractions/hivetracks2023/issues>

2024-02-23

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 interactions dataset of known origin, and discuss their outcome. The dataset under review (aka [globalbioticinteractions/hivetracks2023](https://globalbioticinteractions.org/hivetracks2023)) has size 84.2KiB and contains 1 interaction with 1 unique type of association (e.g., `pollinatedBy`) between 1 primary taxon (e.g., *Trifolium repens*) and 1 associated taxon (e.g., *Apis mellifera ligustica*). The report includes detailed summaries of interactions data as well as a taxonomic review from multiple catalogs.

Table of Contents

Introduction	2
Data Review	2
Methods	2
Results	3
Biotic Interactions	3
Interaction Networks	5
Taxonomic Alignment	6
Additional Reviews	8
GloBI Review Badge	8
GloBI Index Badge	9
Discussion	9
Acknowledgements	10
Author contributions	10

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:

HiveTracks WorldFAIR Test Data.

file:///home/runner/work/hivetracks2023/hivetracks2023/./

For additional metadata related to this dataset, please visit

<https://github.com/globalbioticinteractions/hivetracks2023> 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
elton	0.13.2
nomer	0.5.6
mlr	6.0.0
pandoc	3.1.6.1

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

```
# get versioned copy of the dataset (size approx. 84.2KiB) under review
```

¹ Note that you have to first get the data (e.g., via `elton pull globalbioticinteractions/hivetracks2023`) before being able to generate reviews (e.g., `elton review globalbioticinteractions/hivetracks2023`), extract interaction claims (e.g., `elton interactions globalbioticinteractions/hivetracks2023`), or list taxonomic names (e.g., `elton names globalbioticinteractions/hivetracks2023`)

```
elton pull globalbioticinteractions/hivetracks2023
```

```
# generate review notes
```

```
elton review globalbioticinteractions/hivetracks2023\  
> review.tsv
```

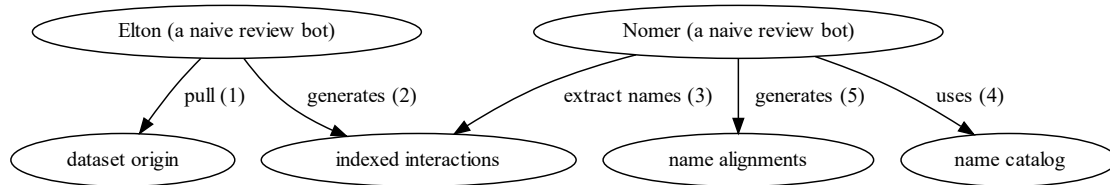
```
# export indexed interaction records
```

```
elton interactions globalbioticinteractions/hivetracks2023\  
> interactions.tsv
```

```
# export names and align them with the Catalogue of Life using Nomer
```

```
elton names globalbioticinteractions/hivetracks2023\  
| nomer append col\  
> name-alignment.tsv
```

or visually, in a process diagram.



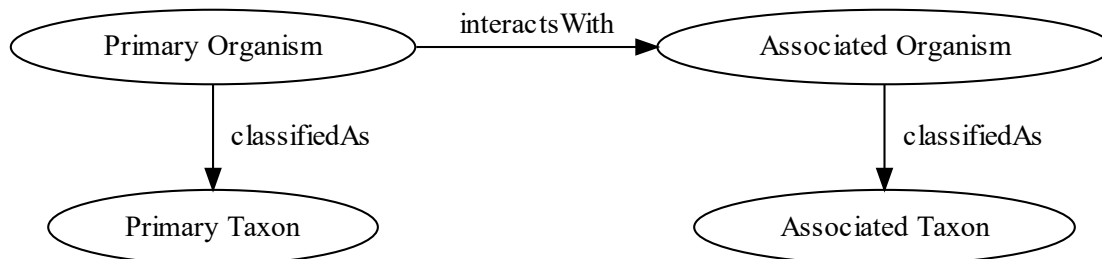
Review Process Overview

You can find a recent copy of the full review script at check-data.sh.

Results

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

Biotic Interactions



Biotic Interaction Data Model

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.

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/hivetracks2023](#)) has size 84.2KiB and contains 1 interaction with 1 unique type of association (e.g., `pollinatedBy`) between 1 primary taxon (e.g., *Trifolium repens*) and 1 associated taxon (e.g., *Apis mellifera ligustica*).

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	interactionTypeNa me	targetTaxonName	referenceCitation
Trifolium repens	pollinatedBy	Apis mellifera ligustica	https://docs.google.com/spreadsheets/d/16DdO55ukMlkfYtJ05FuYSnwrSwNU2QIAAIQT14VF0EM/export?format=tsv&id=16DdO55ukMlkfYtJ05FuYSnwrSwNU2QIAAIQT14VF0EM&gid=776329546

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

interactionTypeName	count
pollinatedBy	1

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

sourceTaxonName	count
Trifolium repens	1

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

targetTaxonName	count
Apis mellifera ligustica	1

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

sourceTaxonName	interactionTypeName	targetTaxonName	count
Trifolium repens	pollinatedBy	Apis mellifera ligustica	1

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](#)



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

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, globi, mdd, tpt, and 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
Apis mellifera ligustica	SYNONYM_OF	col	Apis mellifera
Trifolium repens	HAS_ACCEPTED_NAME	col	Trifolium repens
Apis mellifera ligustica	SYNONYM_OF	discoverlife	Apis mellifera
Trifolium repens	NONE	discoverlife	Trifolium repens

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	species	2
discoverlife	NA	1
discoverlife	species	1
gbif	species	2
gbif	variety	1
globi	NA	1
globi	species	2
globi	subspecies	1
globi	variety	1
itis	species	2
mdd	NA	2
ncbi	species	1
ncbi	subspecies	1
pbdb	NA	2

resolvedCatalogName	resolvedRank	count
tpt	NA	2

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	1
col	HAS_ACCEPTED_NAME	1
discoverlife	SYNONYM_OF	1
discoverlife	NONE	1
gbif	SYNONYM_OF	3
gbif	HAS_ACCEPTED_NAME	3
globi	SAME_AS	21
itis	SYNONYM_OF	1
itis	HAS_ACCEPTED_NAME	1
mdd	NONE	2
ncbi	SAME_AS	2
pbdb	NONE	2
tpt	NONE	2

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)
globi	associated names alignments (first 500, full csv/tsv)

catalog name	alignment results
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.

First few lines in the review notes.

reviewDate	reviewCommentType	reviewComment
2024-02-23T00:37:08Z	note	source taxon name missing
2024-02-23T00:37:08Z	note	source taxon name missing
2024-02-23T00:37:08Z	note	target taxon name missing
2024-02-23T00:37:08Z	note	invalid date string [20231125T000000-0500]

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

Most frequently occurring review notes, if any.

reviewComment	count
source taxon name missing	2
invalid date string [20231125T000000-0500]	2
target taxon name missing	1

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.



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.



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 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-ness⁶ 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.

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

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

⁵ At time of writing (2024-02-23) the version of the GloBI dataset index was available at [\[https://globalbioticinteractions.org/datasets\]](https://globalbioticinteractions.org/datasets)(<https://globalbioticinteractions.org/datasets>)

⁶ 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."

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 means that it can be more precisely processed 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 to facilitate seamless data exchanges, i.e., 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>.