

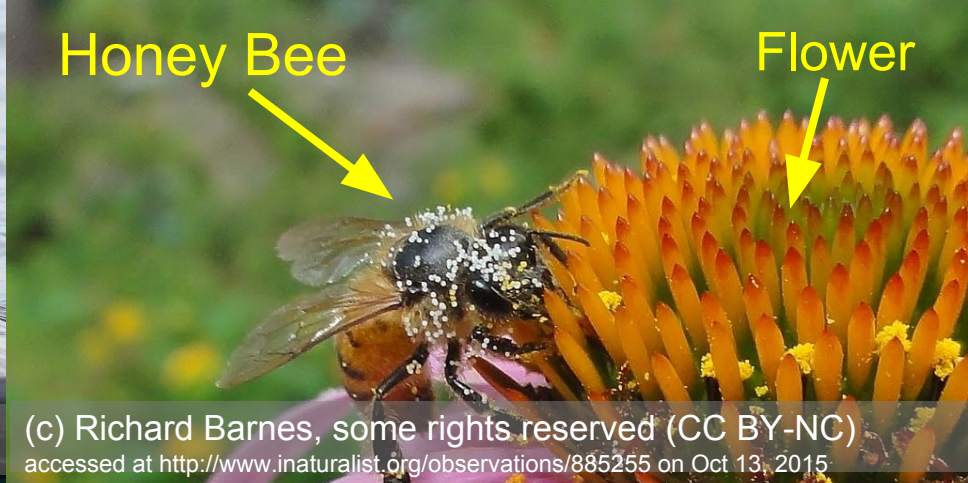
Global Biotic Interactions

A Catalyst for Integrating Existing
Species-Interaction Datasets, Connecting Curators
and Developing Data Exchange Methods

Poelen, J., 2017. Global Biotic Interactions: A Catalyst for Integrating Existing Interaction Datasets, Connecting Data Curators and Developing Data Exchange Methods. Proceedings of TDWG, 1, p.e20214. Available at: <http://dx.doi.org/10.3897/tdwgproceedings.1.20214>.



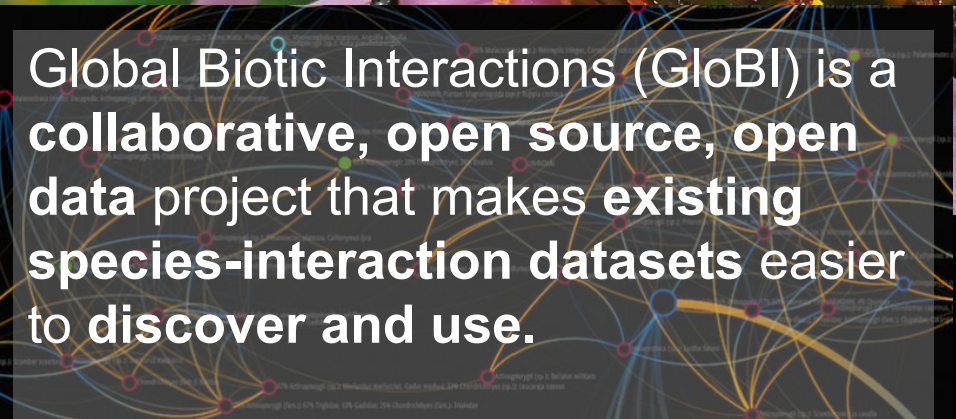
(c) edward_rooks, some rights reserved (CC BY-NC)
accessed at <http://www.inaturalist.org/observations/563486> on Feb 4, 2015



(c) Richard Barnes, some rights reserved (CC BY-NC)
accessed at <http://www.inaturalist.org/observations/885255> on Oct 13, 2015



(c) Cheryl Harleston, some rights reserved (CC BY-NC-SA)
accessed at <http://www.inaturalist.org/observations/2020957> on Oct 13, 2015



Background image: Slyusarev et al. (2015): Global Biotic Interactions food web map. figshare. <http://dx.doi.org/10.6084/m9.figshare.1297762>

<http://globalbioticinteractions.org>

ANIMAL ECOLOGY

BY
CHARLES ELTON

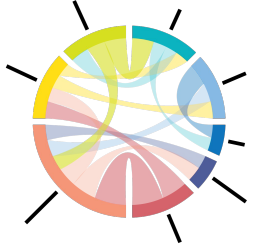
WITH AN INTRODUCTION BY
JULIAN S. HUXLEY, M.A.
FULLERIAN PROFESSOR OF PHYSIOLOGY, ROYAL INSTITUTION

“The advantage, and at the same time the difficulty, of ecological work is that it attempts to provide conceptions which can link up into some complete scheme the colossal store of facts about natural history which has accumulated up to date in this rather haphazard manner. [...] Until more organised information about the subject is available, it is only possible to give a few instances of some of the more clear-cut niches which happen to have been worked out.”

Charles Elton, **1927**, *Animal Ecology*.

NEW YORK
THE MACMILLAN COMPANY

1927



Global Biotic Interactions

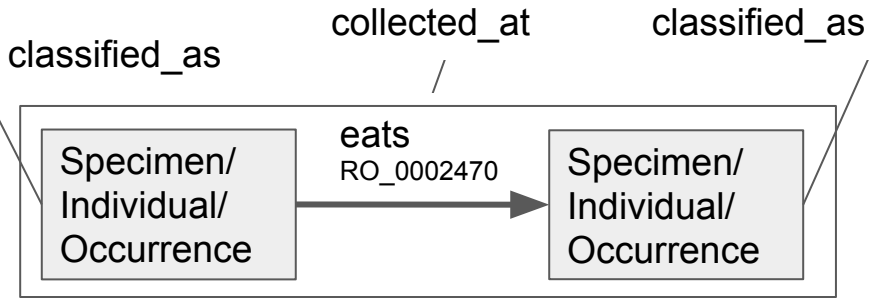
A Catalyst for **Integrating Existing Species-Interaction Datasets**, Connecting Curators and Developing Data Exchange Methods



Southern Sea Otter
(*Enhydra lutris nereis*)
inaturalist.org/taxa/117520

Location
Lat: 36.713851
Lon: -121.960949

Pacific rock crab
(*Romaleon antennarium*)
inaturalist.org/taxa/202315



collected
2014-03-09 PDT

Study
<https://www.inaturalist.org/observations/563486>

Dataset
<https://github.com/globalbioticinteractions/inaturalist/archive/fae51d40d470e9adaa15b89bd2f1ba6c0a5c8fbd.zip>



Simplified internal data model used by GloBI to integrate interaction data.



Ecoregions

Environments

Location
 Lat: 36.713851
 Lon: -121.960949

Pacific rock crab
 (*Romaleon antennarium*)
 inaturalist.org/taxa/202315

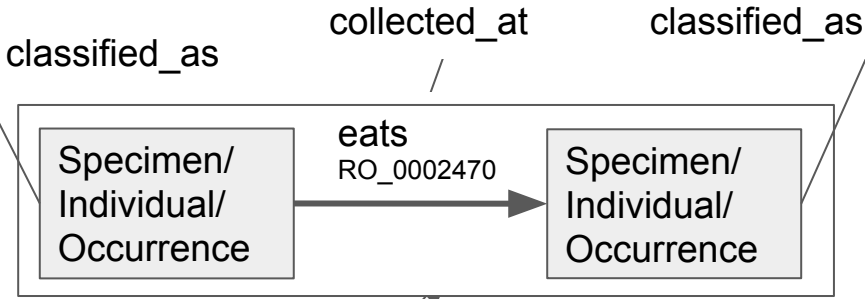
Pacific rock crab
 (*Romaleon antennarium*)
 gbif.org/species/5970317

Pacific rock crab
 (*Romaleon antennarium*)
 marinespecies.org/aphia.php?p=tax
 details&id=440397

Southern Sea Otter
 (*Enhydra lutris nereis*)
 inaturalist.org/taxa/117520

Southern Sea Otter
 (*Enhydra lutris nereis*)
 marinespecies.org/aphia.php?p
 =taxdetails&id=242601

Southern Sea Otter
 (*Enhydra lutris nereis*)
 gbif.org/species/6163936



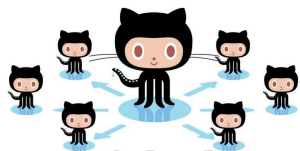
Study
<https://www.inaturalist.org/observations/563486>

Dataset
<https://github.com/globalbioticinteractions/inaturalist/archive/fae51d40d470e9adaa15b89bd2f1ba6c0a5c8fbd.zip>



Simplified internal data model used by GloBI to integrate interaction data.

federated
data registries



github
SOCIAL CODING
data = code

discover, import,
aggregate and link



open
read-only
archives



open
read-only
search



{Web API}
{rglobi}
{globi-js}



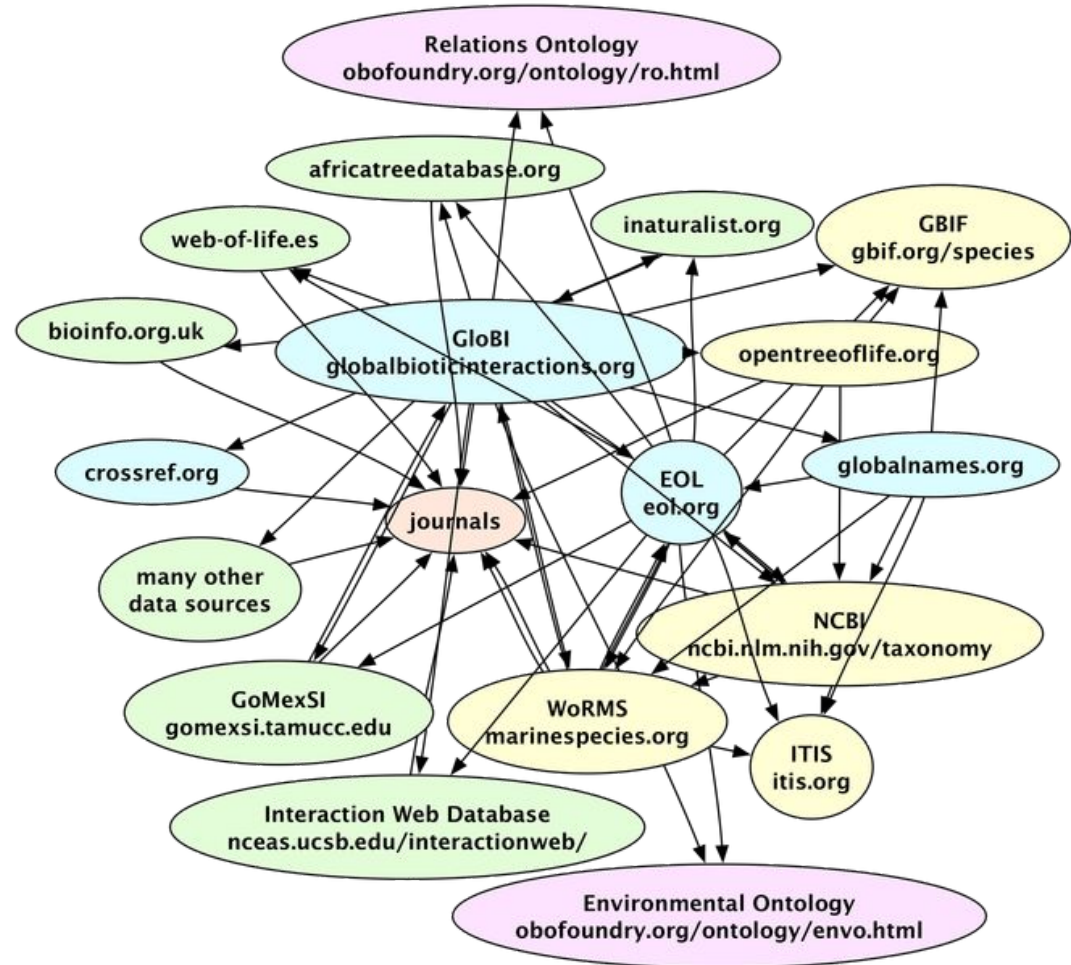
an automated and continuous process



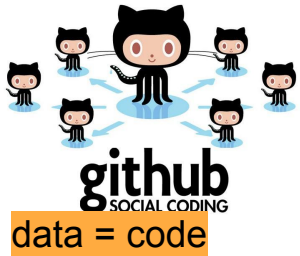
Bidirectional links include Encyclopedia of Life, Gulf of Mexico Species Interactions, NCBI Taxonomy, World Register of Marine Species, iNaturalist, Fishbase and SeaLifeBase.

Outgoing links include UBERON (body parts, life stage, physiological state), EnvO, GeoNames, CMECS, FEOW, MEOW, TEOW, doi.org, ITIS, Open Tree of Life, NBN and ALA.

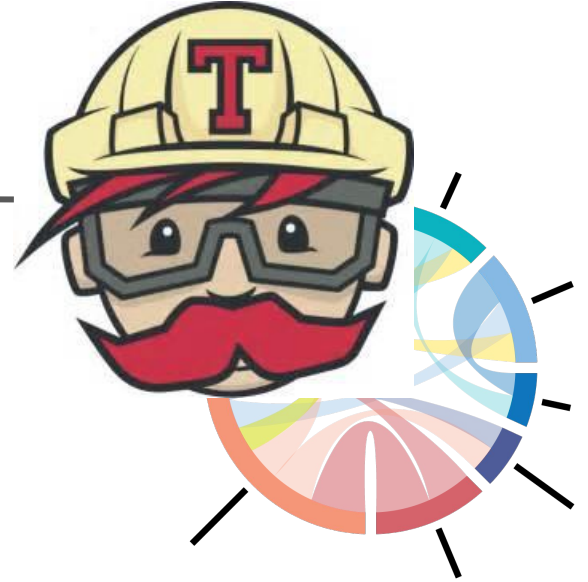
Link services include Global Names and CrossRef.



federated
data registries



peer review



an automated and continuous process



Lifestages of Species Interaction Datasets

Jan 24, 2017



Just like organisms, datasets get born, grow up, reproduce, and die. One of our main goals is to help increase the productivity (or reuse), and increase the lifespan of datasets.

To monitor the lifestage of a dataset, the GloBI "status" page was introduced in January 2017. This page shows the state of the featured collection of species interaction datasets that make up GloBI. Rather than treating datasets like static entities, GloBI takes a dynamic approach and treats datasets organically to incorporate change and reward the links to naming authorities (e.g. tax.gov), publishers (e.g. publishers.org) and other relevant external data services. Dataset numbers can be found on the status page that indicate whether the dataset can be read, searched or cited. In addition, some statistics are provided to point to known issues and the challenges of the dataset (e.g. number of interactions, number of names).

Each row of names). With this, a de-facto publication process is taking along with quality control measures that show how the lifestage of a dataset.

At a time of writing, the status page provides a wealth of information. For instance, the state of the dataset (e.g. **readable**) provided by <http://highlights.ccc.edu> (GloBI-Liu et al. 2016) indicates that the dataset is close to being in already read: many of its interactions are not and unprocessed issues exist. In contrast, the chance of survival of a dataset like the Africa Tree Database (Gardner et al. 2015) looks promising: interactions are green (100%), name match rate is 94% across 1.3k names and 7.7k interactions. In addition, the Africa Tree Database has been deposited with Zenodo, a service that is integral to greater "permanent" data availability through digital object identifiers.

Over the last couple of weeks, I've used the page to discover and resolve various dataset issues. For example, an obscure GloBI tag was being used (resolved) (see <http://bioticinteractions.org/2017/01/24/lifestages-of-species-interaction-datasets/>) and <http://interactions.globi-dataset.org/> had prevented the integration of some interaction records provided by the Africa Tree Database. I am curious to see how the status page will evolve in the next months.

For more information about the status page, please visit the [link](#).

management site page. The development of the status page was supported by the Encyclopedia of Life.



Just like organisms, datasets get born, grow up, reproduce and die. GloBI's mission is to help increase the productivity (or reuse) and lifespan of datasets before they meet their maker.

readable issues search citable stats

build	unknown	issues	0 open	GloBI	DOI	stats	dataset name
build	unknown	issues	0 open	GloBI ✓	DOI ✗	9.1k / 4.3k / 87% / 6d	cmungall/Benesh-et-al-2017
build	unknown	issues	0 open	GloBI ✓	DOI ✗	16 / 25 / 100% / 6d	cmungall/dinosaur-biotic-interactions
build	passing	issues	0 open	GloBI ✓	DOI ✗	331 / 138 / 98% / 6d	diatomsRcool/greenland_interactions
build	passing	issues	0 open	GloBI ✓	DOI ✗	254 / 252 / 99% / 6d	diatomsRcool/yellowstone_grizzly
build	unknown	issues	0 open	GloBI ✗	DOI ✗	6d	Dryu0003/dietdatabase
build	unknown	issues	2 open	GloBI ✓	DOI ✗	183.8k / 34.2k / 99% / 6d	EOL/pseudonitzchia

GloBI currently includes **268,317** references obtained from **310** data sources. In total, **2,852,596** interaction records were discovered, covering **181,570** taxa. A [taxon map](#) shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching algorithm are documented in the [list of unmatched taxon names by reference/source](#). These unmatched or unresolved names are typically unknown or invalid names.

Below, you can search for references that contain species interaction records. Example queries: Which references document sea otters (*Enhydra lutris*) prey? or Who documented what honey bees (*Apis*) pollinate?

Which references containing claim that interacts with ?

<http://globalbioticinteractions.org/references>

Accessed at 28 Sept 2018

<https://gbif.org>

Accessed at 28 Sept 2017

2.8M records

0.1k datasets

~100k **taxa**

950.6M records

36.5k datasets

~1-2M **species**

Eltonian shortfall*: a lack of species-interaction records

*Hortal, J. et al., 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46(1). Available at: <http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400>.

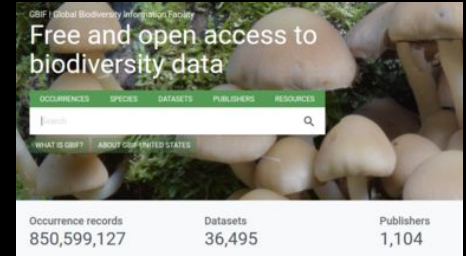
GloBI currently includes **293,203 references** obtained from **326 data sources**. In total, **3,379,426 interaction records** were discovered, covering **233,557 taxa**. A **taxon map** shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching algorithm are documented in the list of **unmatched taxon names by reference/source**. These unmatched or unresolved names are typically unknown or invalid names.

Below, you can search for references that contain species interaction records. Example queries: *Which references document sea otters (*Enhydra lutris*) prey?* or *Who documented what honey bees (*Apis*) pollinate?*

Which references containing claim that ?

<http://globalbioticinteractions.org/references>

Accessed at 26 Feb 2018



Free and open access to biodiversity data

Occurrences: 850,599,127 | Datasets: 36,495 | Publishers: 1,104

<https://gbif.org>

Accessed at 26 Feb 2018

3.4M records

0.1k datasets

~100k taxa

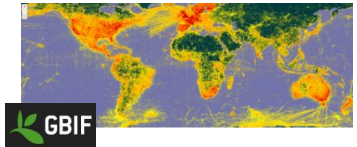
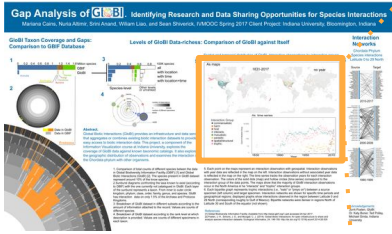
972.7M records

38.1k datasets

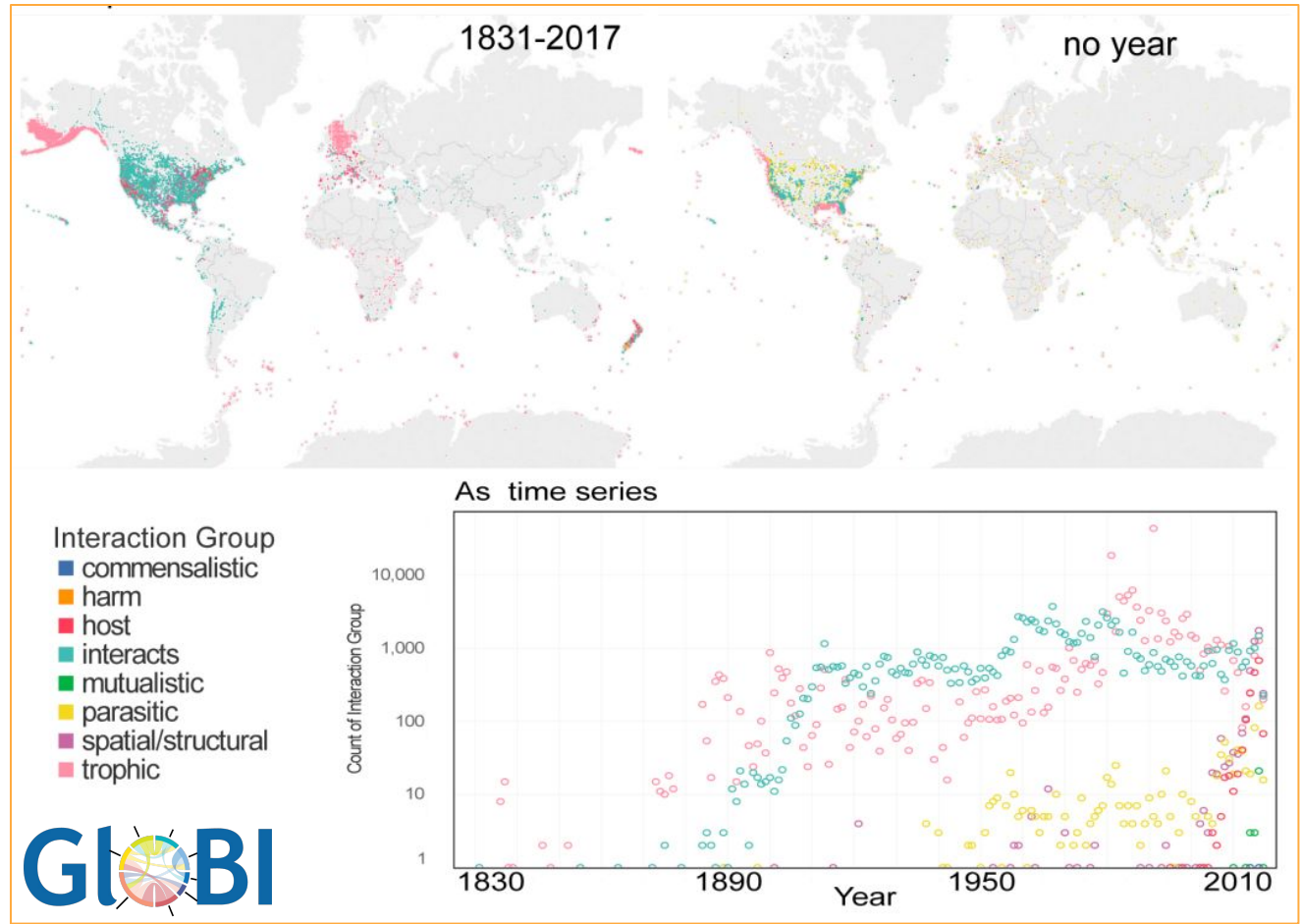
~1-2M species

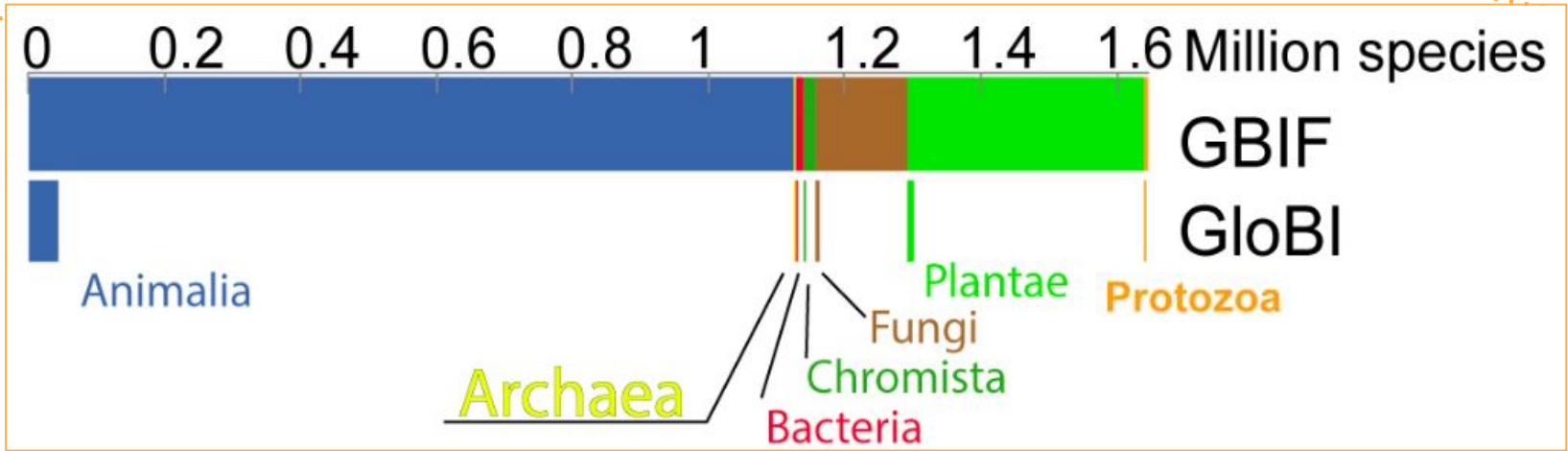
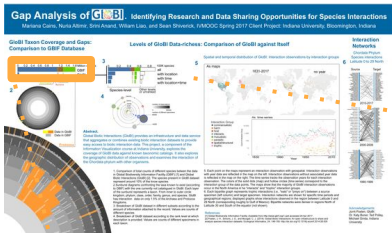
Eltonian shortfall*: a lack of species-interaction records

*Hortal, J. et al., 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46(1). Available at: <http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400>.

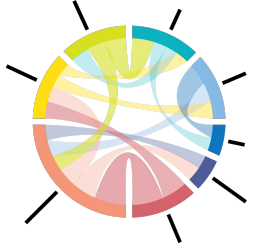


geospatial
and
temporal
Eltonian
shortfall



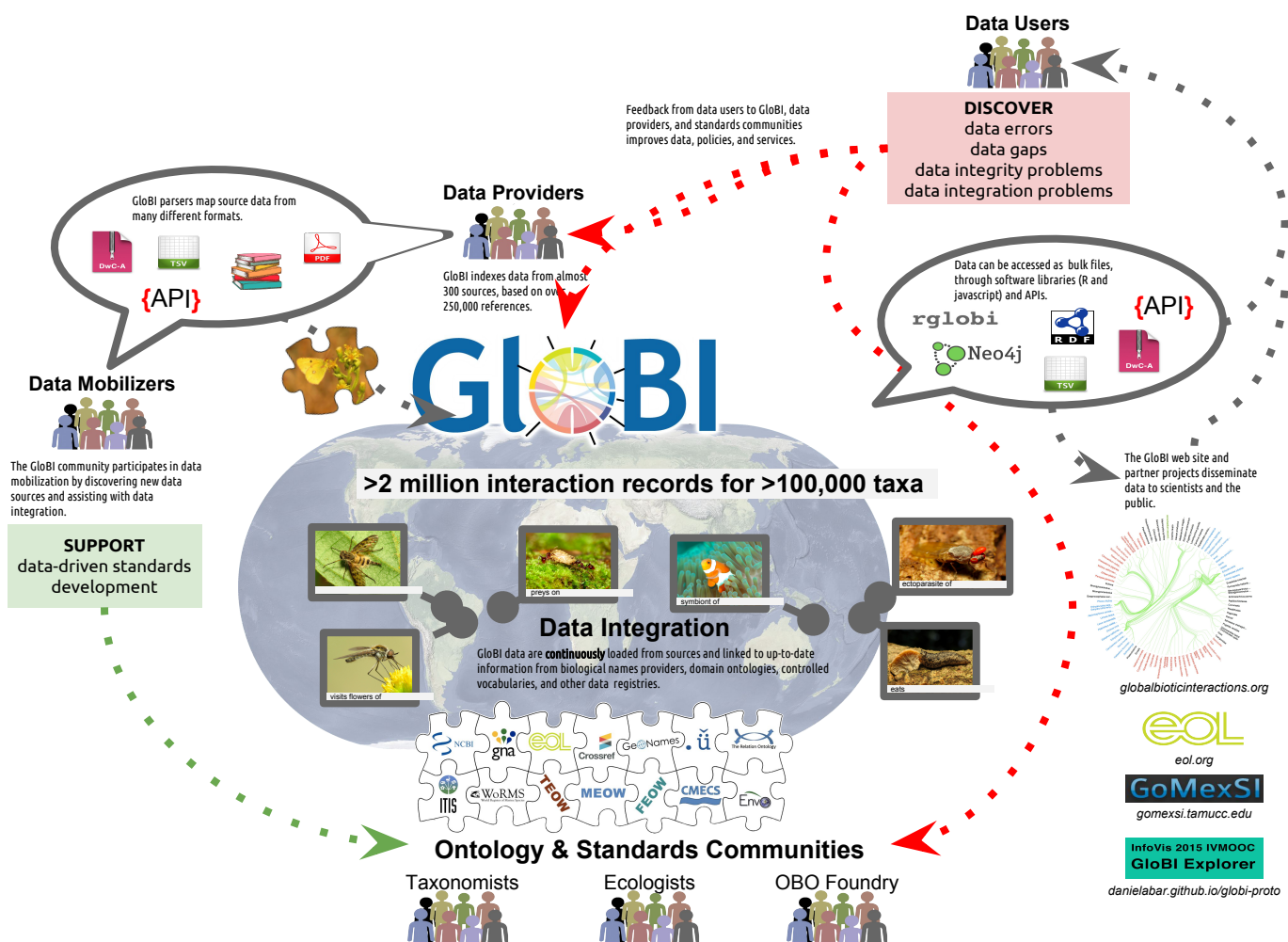


taxonomic Eltonian shortfall

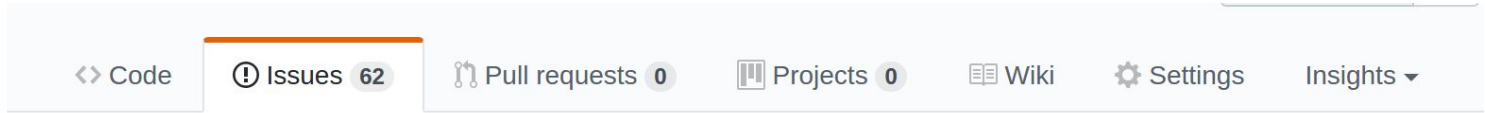


Global Biotic Interactions

A Catalyst for Integrating Existing
Species-Interaction Datasets, **Connecting Curators**
and Developing Data Exchange Methods



Open discussions and reviews using GitHub issues



A life cycle database for parasitic acanthocephalans, cestodes, and nematodes by Benesh et al. #305

 **Open** jhpoelen opened this issue on Aug 14 · 16 comments



jhpoelen commented on Aug 14 • edited

Owner



As suggested by [@derele](#) -

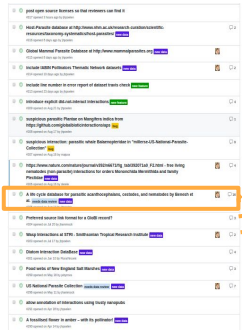
[...]

I am currently interacting with Dan Benesh who compiled a database on parasite life cycles (<http://onlinelibrary.wiley.com/doi/10.1002/ecy.1680/full>). Dan will work with my team in Berlin soon. He told me that he also wants to include this data in globi,

[...]

<https://github.com/jhpoelen/eol-globi-data/issues/305> accessed on 28 Sept 2017

Open discussions and reviews using GitHub issues



(<http://onlinelibrary.wiley.com/doi/10.1002/ecy.1680/full>) . Dan will work with my team in Berlin soon. He told me that he also wants to include this data in globi, [...]

 **millerse** self-assigned this on Aug 21

 **millerse** added the **new data** label on Aug 21



dbenesh82 commented 28 days ago



Hi, I compiled this parasite life cycle database and have been meaning to get it into globi. If anything is unclear, I would be happy to help.



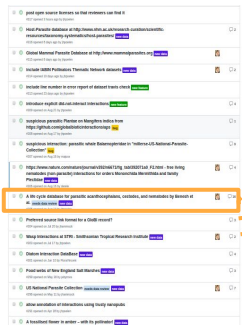
millerse commented 9 days ago

Collaborator



And done [see here](#)

Open discussions and reviews using GitHub issues



millerse commented 9 days ago

Collaborator



And done [see here](#)



millerse added the **needs data review** label 9 days ago



jhammock commented 9 days ago

Collaborator



Looks good! If you want to get fancy, you will probably find terms for some of the LocationinhostName values in [UBERON](#). I wouldn't bother chasing elusive ones, but if you find an exact match to the string you have, that should be worth using.



cmungall commented 9 days ago

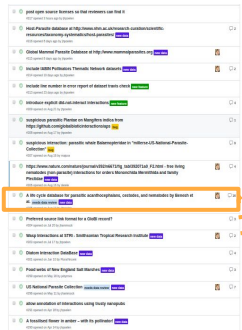
Collaborator



Scanned the 78 distinct values, fairly sure we have them all, I can do the mapping if you like



Open discussions and reviews using GitHub issues



jhammock commented 9 days ago

Collaborator



Fantastic! Yes please



cmungall commented 9 days ago

Collaborator



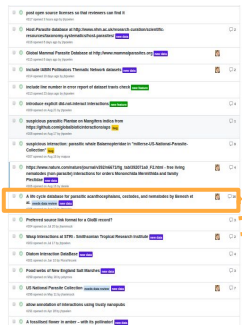
consider also mapping these to ENVO:

```
Host.habitat: freshwater
Host.habitat: marine
Host.habitat: terrestrial
```

These to PATO:

```
Shape: NA
Shape: coiled
Shape: cylinder
Shape: ellipsoid
```

Open discussions and reviews using GitHub issues



jhammock commented 6 days ago

Collaborator



using endoparasite is easy enough. **@dbenesh82**, can you verify that the definition and context displayed at http://purl.obolibrary.org/obo/RO_0002634 meet your needs? I just came from an ontology workshop and have discovered how easily definitions can be found incompatible...



dbenesh82 commented 4 days ago



Yes **@jhammock**, endoparasite is more appropriate for the interactions in this dataset.



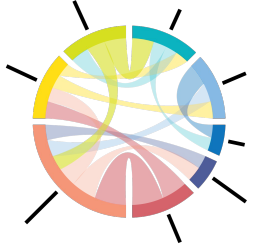
millerse commented 2 days ago

Collaborator



Okay everyone. I have added the endoparasite link and the body terms to the sheet.

@dbenesh82 To your comment about the extra numbers: You had multiple body locations in a single record (i.e. gut, body cavity, etc.). I took each of those terms and made a single full record for each term. That is where the records come from.



Global Biotic Interactions

A Catalyst for Integrating Existing
Species-Interaction Datasets, Connecting Curators
and **Developing Data Exchange Methods**



Supports ~40 existing species-interaction data formats
Exports rdf, neo4j, tsv and darwin core (-ish) archives
Integrates with various ontologies and taxonomies

Rebuilt from sources automatically and continuously
Offline workflows via elton, elton-archive and nomer*

Offers open source and open access data
Encourages to share structured datasets
Provides a community hub for data curators and users

*see <https://github.com/globalbioticinteractions/>



Acknowledgments / funding

an incomplete list in no particular order

GloBI is not possible without the many contributions (big and small) of folks like Jen Hammock, Katja Schulz, Pepper Luboff, Chris Mungall, Katja Seltmann, Brian Hayden, Ken-ichi Ueda, Mariana Cains, Nuria Altimir, Srinu Anand, William Liao, Sean Shiverick, Jim Simons, Theresa Mitchell, Emanuel Heitlinger, Marius Bäsler, Kathy Kwan, Deng Palomares, Josephine “Skit” Barile, Anne Thessen, Allen Hurlbert, Malcolm Storey... and thousands of others that have collected and shared species-interaction data.

GloBI has received funding from several projects since 2013. Those funding sources include, but are not limited to, the Encyclopedia of Life, EOL Rubenstein Fellows Program (CRDF EOL-33066-13/F33066, 2013) and the David M. Rubenstein Grant (FOCX-14-60988-1, 2014), and the Smithsonian Institution (SI) (T15CC10297-002, 2016).



<https://globalbioticinteractions.org>
info@globalbioticinteractions.org

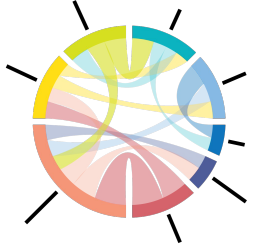
Please cite GloBI using;

Jorrit H. Poelen, James D. Simons and Chris J. Mungall. (2014).
Global Biotic Interactions: An open infrastructure to share and analyze
species-interaction datasets. *Ecological Informatics*.
<http://dx.doi.org/10.1016/j.ecoinf.2014.08.005>.







Workshop Prompt

What can you do to help your colleagues and friends to share interactions datasets and make sure that they **remain accessible** (even after GloBI ceases to exist)?



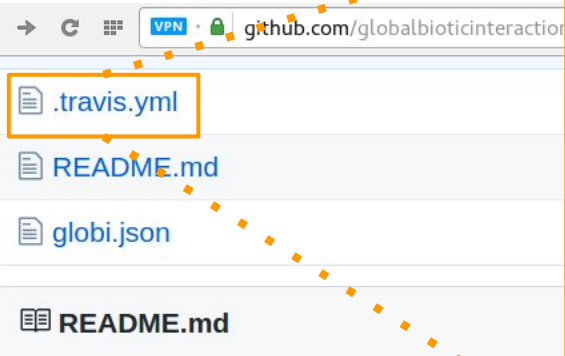
Extra

 .travis.yml	add travis and badges	9 months
 README.md	update dataset doi	8 months
 globi.json	adding null value mapping	2 years
 README.md		

AfricaTreeDatabase

build passing DOI [10.5281/zenodo.229547](https://doi.org/10.5281/zenodo.229547) GloBI ✓

Seltzer, Carrie; Wysocki, William; Palacios, Melissa; Eickhoff, Anna; Pilla, Hannah; Aungst, Jordan; Mercer, Aaron; Quicho, Jamie; Voss, Neil; Xu, Man; J. Ndangalasi, Henry; C. Lovett, Jon; J. Cordeiro, Norbert (2015): Plant-animal interactions from Africa. figshare. <https://dx.doi.org/10.6084/m9.figshare.1526128> . To be included in <http://globalbioticinteractions.org> .



```
install:
```

```
- wget "https://raw.githubusercontent.com"
- chmod +x check-dataset.sh
```

```
script:
```

```
- ./check-dataset.sh ${TRAVIS_REPO_SLUG}
```

AfricaTreeDatabase

build passing

DOI [10.5281/zenodo.229547](https://doi.org/10.5281/zenodo.229547)

GloBI ✓

Seltzer, Carrie; Wysocki, William; Palacios, Melissa; Eickhoff, Anna; Pilla, Hannah; Aungst, Jordan; Mercer, Aaron; Quicho, Jamie; Voss, Neil; Xu, Man; J. Ndangalasi, Henry; C. Lovett, Jon; J. Cordeiro, Norbert (2015): Plant-animal interactions from Africa. figshare. <https://dx.doi.org/10.6084/m9.figshare.1526128> . To be included in <http://globalbioticinteractions.org> .