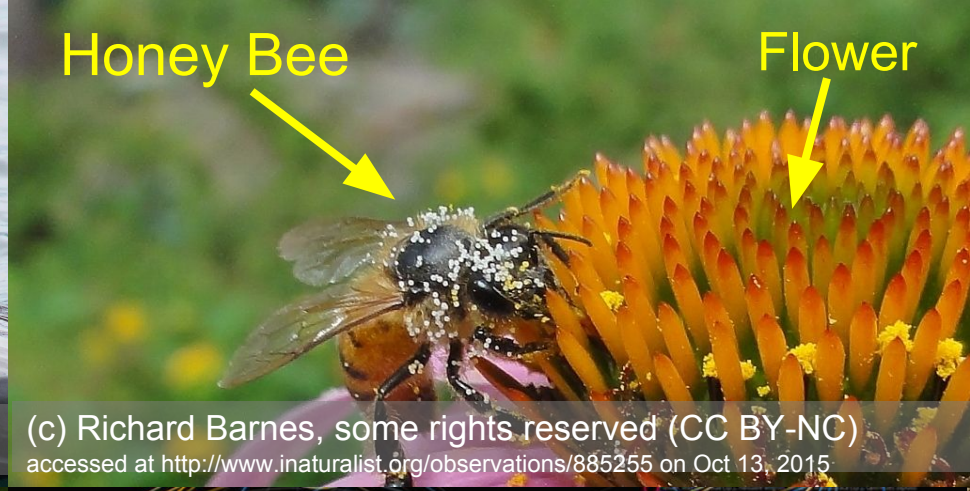
Global Biotic Interactions

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

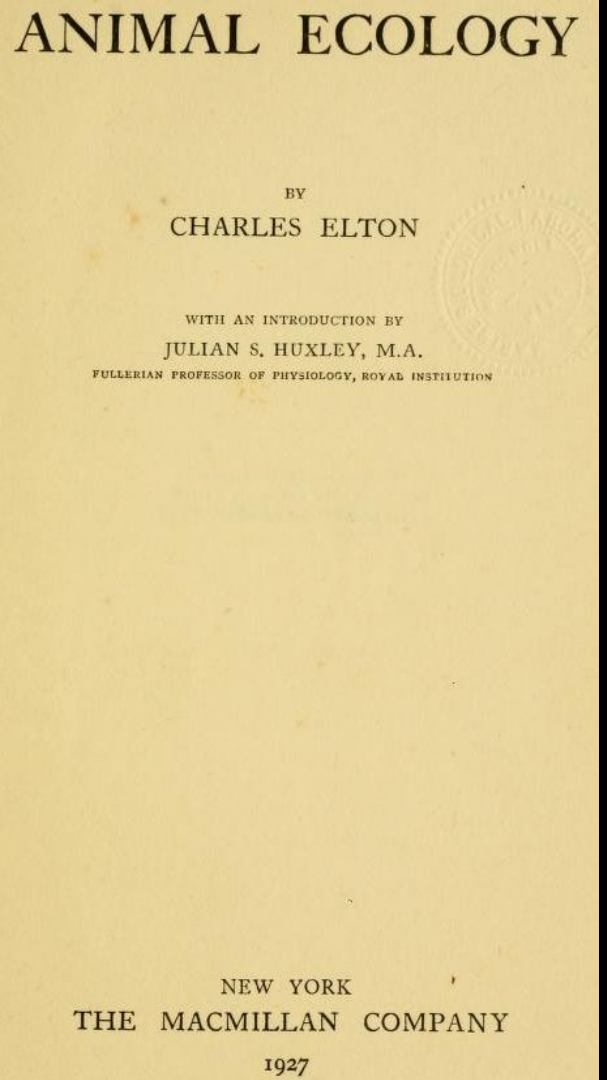
Poelen, J., 2017. Global Biotic Interactions: A Catalyst for Integrating Existing Interaction Datasets, Connecting Data Curators and Developing Data Exchange Methods. Proceedings of TDWG, 1, p.e20214. Available at: http://dx.doi.org/10.3897/tdwgproceedings.1.20214.



Global Biotic Interactions (GloBI) is a **collaborative, open source, open data** project that makes **existing species-interaction datasets** easier to **discover and use.**

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



“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*.

Global Biotic Interactions

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

Location
Lat: 36.713851
Lon: -121.960949

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

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

classified_as

collected_at

classified_as

Specimen/ Individual/ Occurrence

eats
RO_0002470

Specimen/ Individual/ Occurrence

collected
2014-03-09 PDT

in_dataset

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

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

GloBI

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=taxdetails&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

classified_as

collected_at

classified_as

Specimen/ Individual/ Occurrence

eats
RO_0002470

Specimen/ Individual/ Occurrence

collected
2014-03-09 PDT

in_dataset

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

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

GloBI

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

federated
data registries

discover, import,
aggregate and link

open
read-only
archives

zenodo

GloBI

INTERNET ARCHIVE

S3

github
SOCIAL CODING

data = code

neo4j

open
read-only
search

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

Jena

an automated and continuous process

links

GloBI

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.

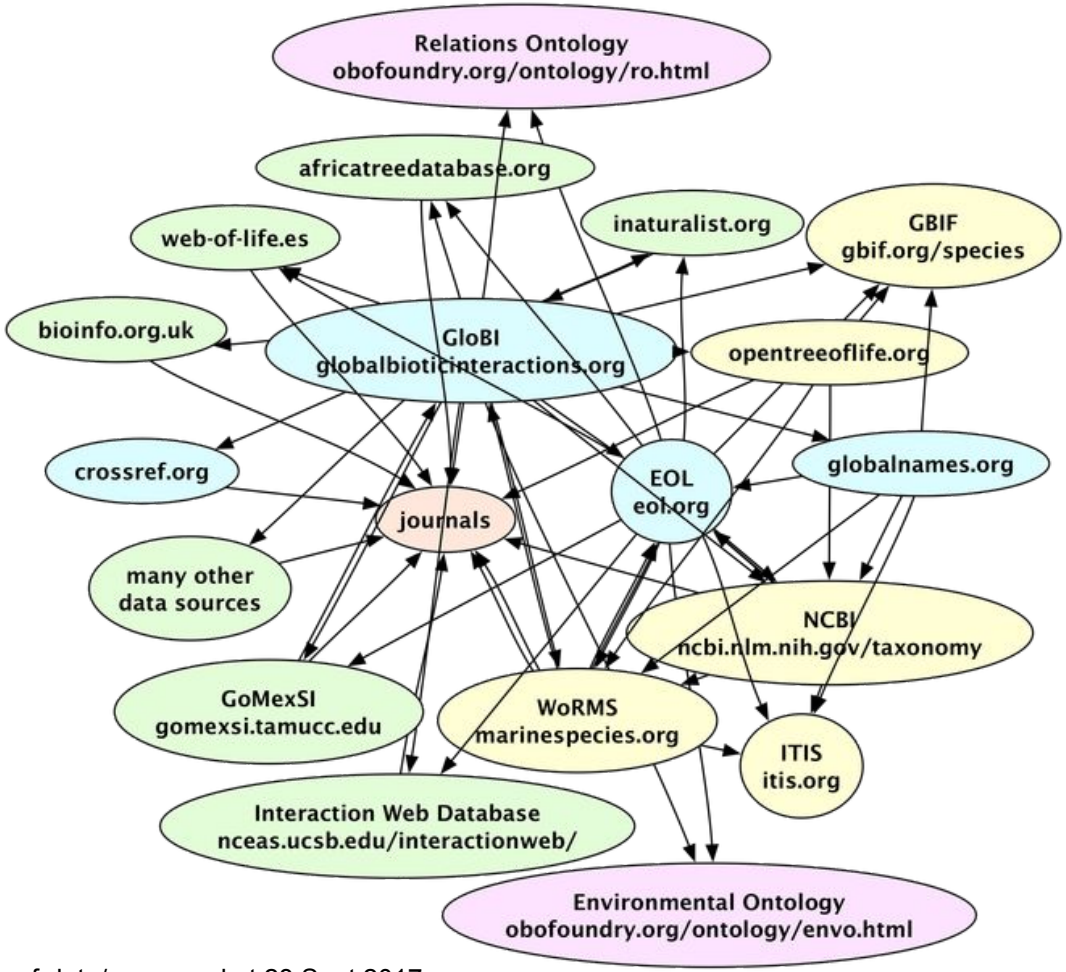
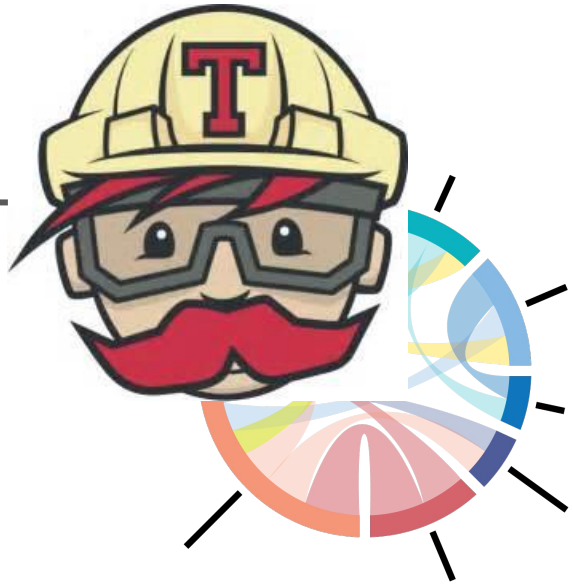


Figure from http://globalbioticinteractions.org/2016/02/08/ecology-of-data/ accessed at 28 Sept 2017

federated
data registries

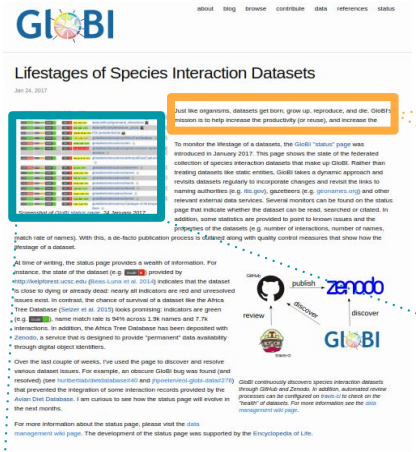
data = code

peer review



an automated and continuous process

http:/globalbioticinteractions.org/2017/01/24/lifestages-of-species-interaction-datasets/



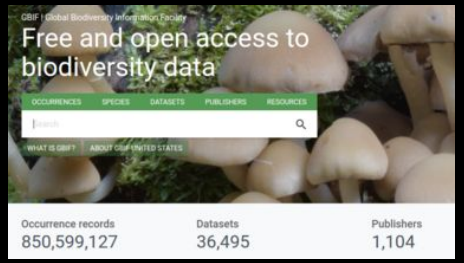
Just like organisms, datasets get born, grow up, reproduce and die. GloBI's mission is to help increase the productivity (or reuse) and lifespan of datasets before they meet their maker.

readable	issues	search citable		stats	
build unknown	issues 0 open	GloBI ✓	DOI ✖	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

http://www.globalbioticinteractions.org/status.html accessed at 28 Sept 2017

http://globalbioticinteractions.org/references
Accessed at 28 Sept 2018

GBIF Global Biodiversity Information Facility



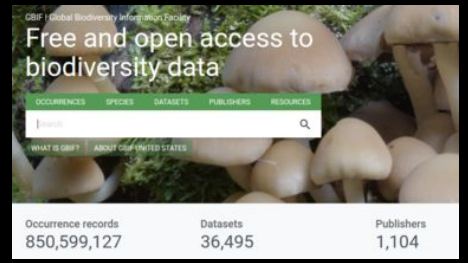
https://gbif.org
Accessed at 28 Sept 2017

2.8M records	950.6M records
0.1k datasets	36.5k datasets
~100k **taxa**	~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.

http://globalbioticinteractions.org/references
Accessed at 26 Feb 2018

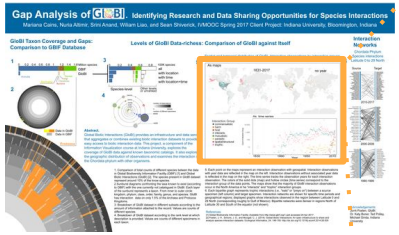


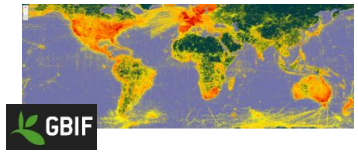
https://gbif.org
Accessed at 26 Feb 2018

3.4M records	**972.7M records**
0.1k datasets	38.1k datasets
~100k **taxa**	~1-2M **species**

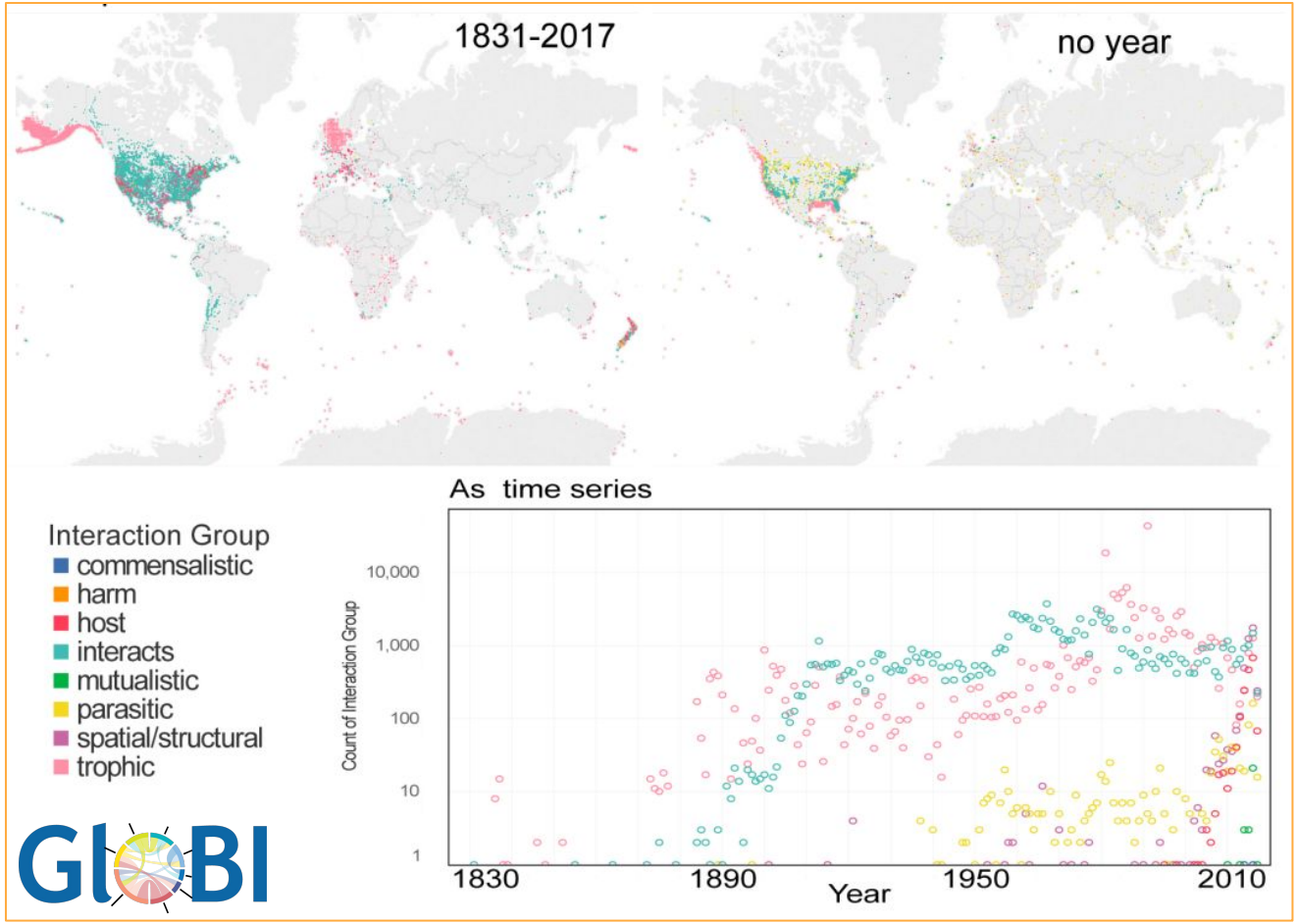
Eltonian shortfall*: a lack of species-interaction records

*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.

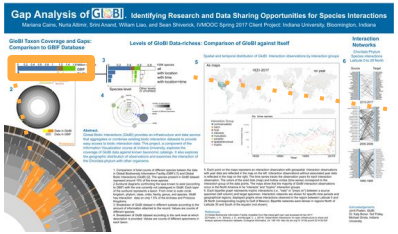


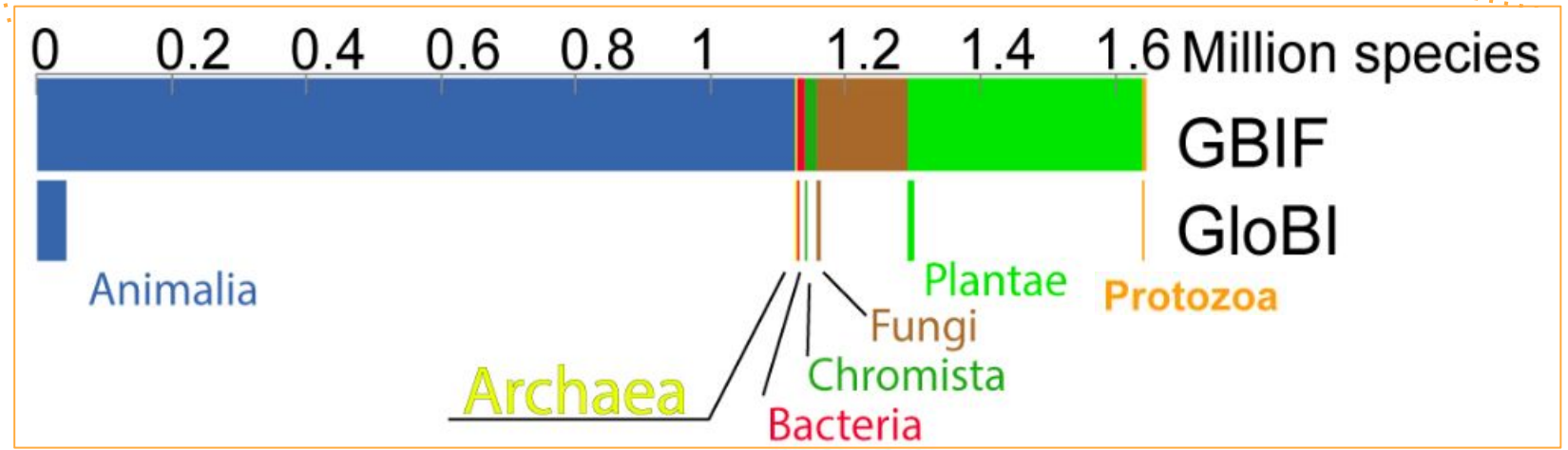


geospatial and temporal Eltonian shortfall



Cains, Mariana, Altimir, Nuria, Anand, Srini, Liao, William, & Shiverick, Sean. (2017). IVMOOC 2017 - Gap Analysis of GloBI Visualizations. Zenodo. http://doi.org/10.5281/zenodo.814922



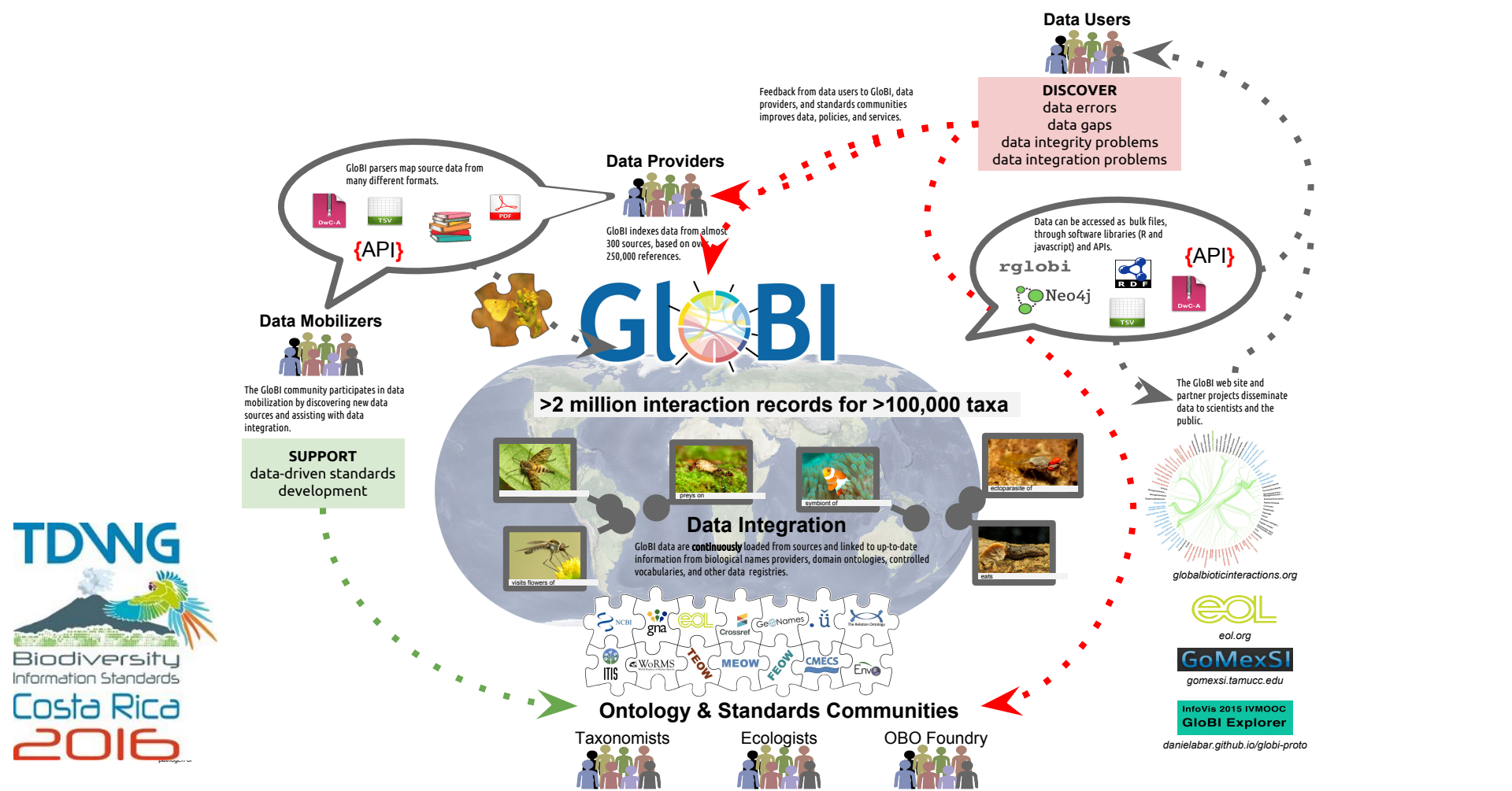


taxonomic Eltonian shortfall

Cains, Mariana, Altimir, Nuria, Anand, Srini, Liao, William, & Shiverick, Sean. (2017). IVMOOC 2017 - Gap Analysis of GloBI Visualizations. Zenodo. http://doi.org/10.5281/zenodo.814922

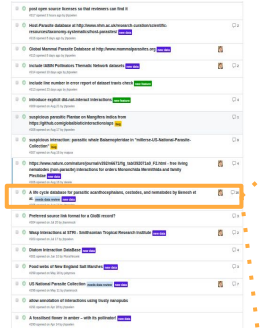
Global Biotic Interactions

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



Poelen JH, Schulz K and Hammock J., 2016. Pragmatic, scalable aggregation of organismal interaction data. Poster presented at the Annual Conference of TDWG in Santa Clara de San Carlos, Costa Rica.

Open discussions and reviews using GitHub issues



Code | Issues 62 | Pull requests 0 | Projects 0 | Wiki | Settings | Insights

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

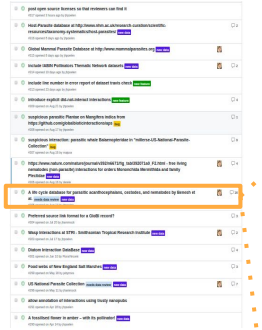
jhpoelen commented on Aug 14 • edited — Owner

As suggested by **@derele** -

[...]
I a 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

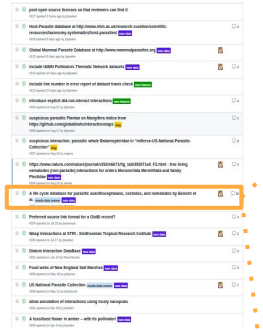
dbenesh82 commented 28 days ago

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

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



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

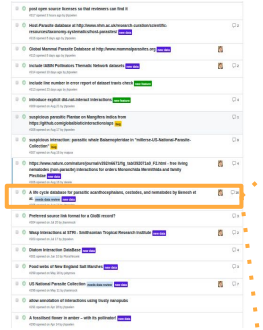
Looks good! If you want to get fancy, you will probably find terms for some of the LocationinhostName values in UBERON. I wouldn't bother chasing elusive ones, but if you find an exact match to the string you have, that should be worth using.

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

...

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



Open discussions and reviews using GitHub issues

jhammock commented 9 days ago — Collaborator

Fantastic! Yes please

cmungall commented 9 days ago — Collaborator

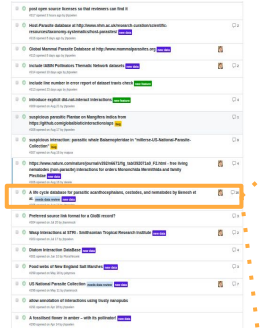
consider also mapping these to ENVO:

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

These to PATO:

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

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



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.

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

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, Srini 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

github.com/globalbioticinteractions/AfricaTreeDatabase

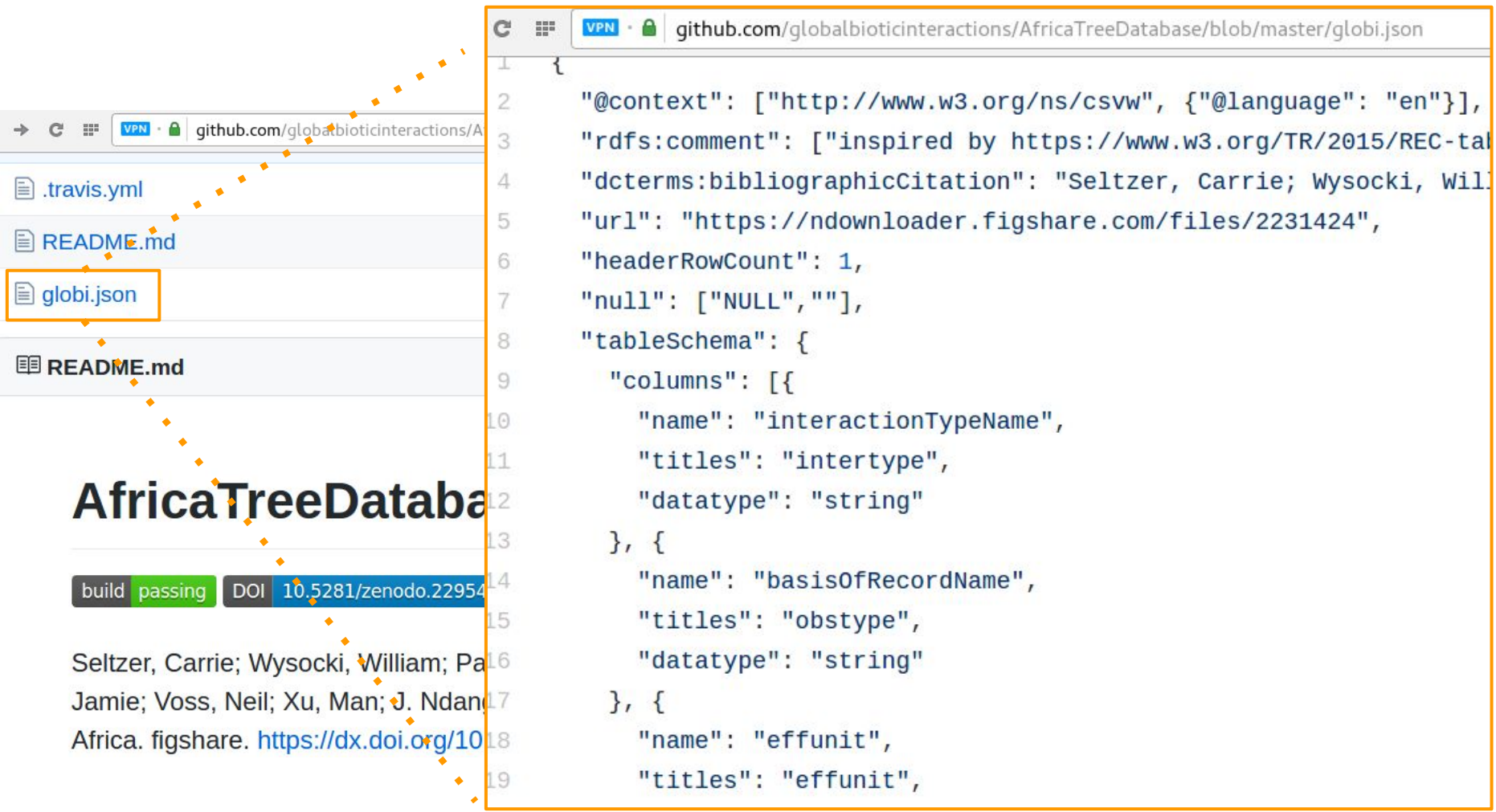
.travis.yml	add travis and badges	9 mont
README.md	update dataset doi	8 mont
globi.json	adding null value mapping	2 yea

README.md

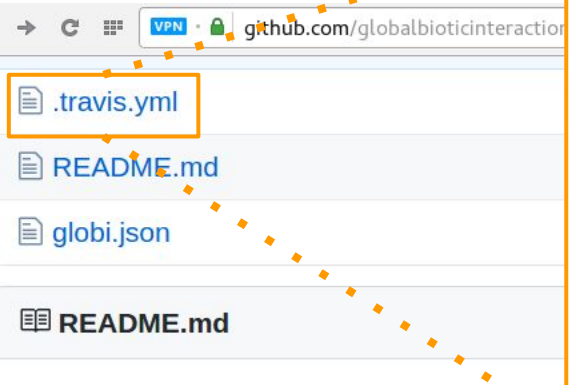
AfricaTreeDatabase

build passing | DOI 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 .



Model for Tabular Data and Metadata on the Web http://w3.org/TR/2015/REC-tabular-data-model-20151217/




```
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 | 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 .

Model for Tabular Data and Metadata on the Web http://w3.org/TR/2015/REC-tabular-data-model-20151217/