



Interaction Data Management, Reuse and Publication

Jorrit H. Poelen

Dead Wood Workshop 2021-10-25



Jorrit H. Poelen is an independent software/data engineer, Ronin research scholar, and UCSB CCBER research affiliate with an interest in open science and biodiversity informatics. He lives and works in Oakland, CA.

He contributes to Global Biotic Interactions, Nomer, Preston, Elton, Terrestrial Parasite Tracker, and the Open Traits Network.

He also writes peer-reviewed papers, gives talks and reviews data publications.

For more information, please have a look at Jorrit's CV, a 2020 NSF Biosketch, or a 2020 NIH Biosketch.

Why reuse biodiversity data?

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*.

Center for Advancement and Synthesis of Open Environmental Data and Sciences

PROGRAM SOLICITATION

NSF 21-549



National Science Foundation

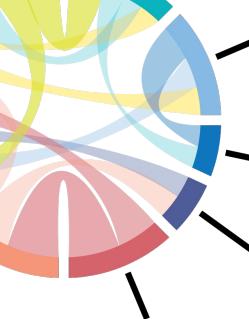
Directorate for Biological Sciences
Division of Biological Infrastructure

Directorate for Computer and Information
Science and Engineering
Office of Advanced Cyberinfrastructure

Letter of Intent Due Date(s) (required) (due by 5 p.m.
submitter's local time):

April 01, 2021

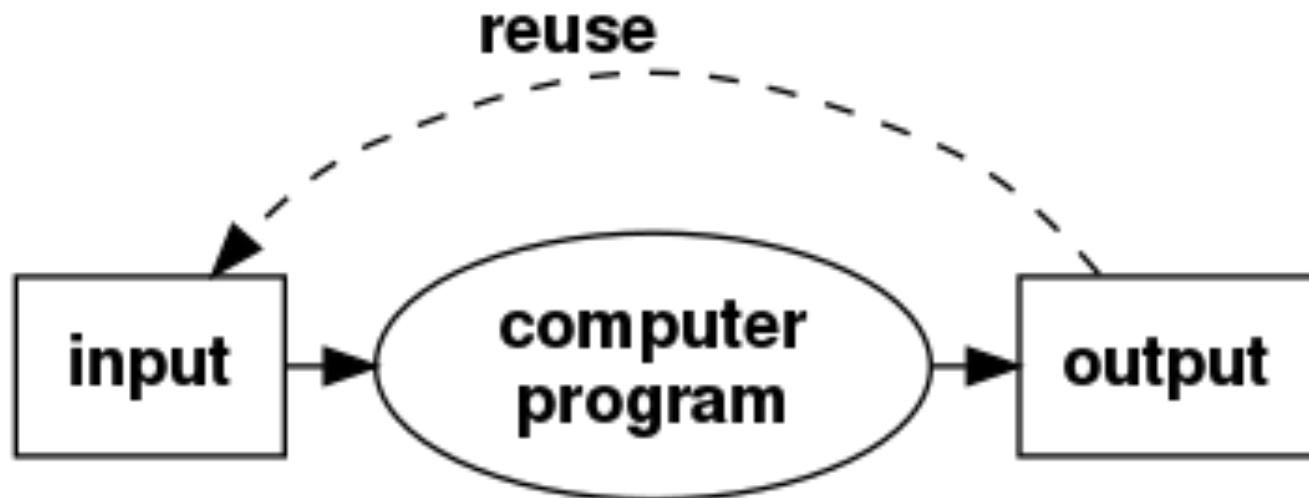
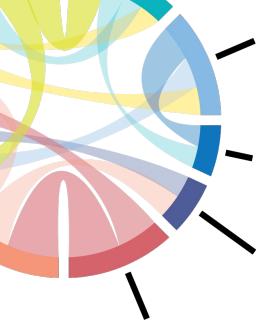
"NSF seeks to establish a Center fueled by open and freely available biological and other environmental data to catalyze novel scientific questions in environmental biology through the use of data-intensive approaches, team science and research networks, and training in the accession, management, analysis, visualization, and synthesis of large data sets. [...] The Center will provide vision for speeding discovery through the increased use of large, publicly accessible datasets to address biological research questions through collaborations with scientists in other related disciplines. [...]"

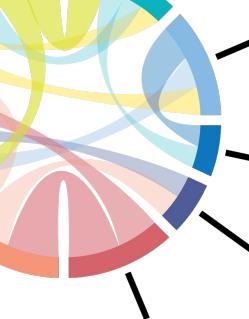


Unix philosophy

Expect the output of every program
to become the input to another, as
yet unknown, program.

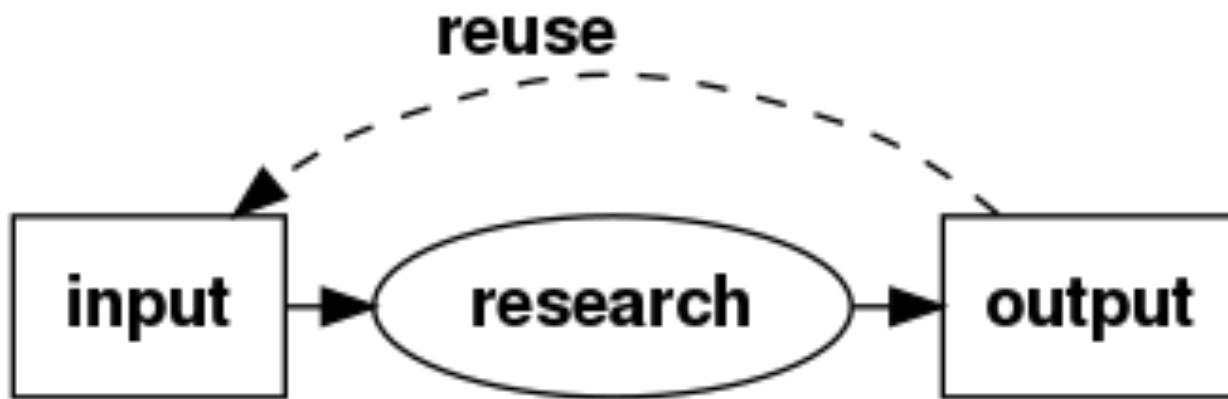
Doug McIlroy, E. N. Pinson, B. A. Tague (8 July 1978). "Unix Time-Sharing System: Foreword". The Bell System Technical Journal. Bell Laboratories. pp. 1902–1903.





a research philosophy

Expect the output of every ~~program~~
research project to become the
input to another, as yet unknown,
~~program~~ **research project.**



How to compile, review, publish, and (re)use big and complex datasets?

automated workflow /
version data /
continuous review /



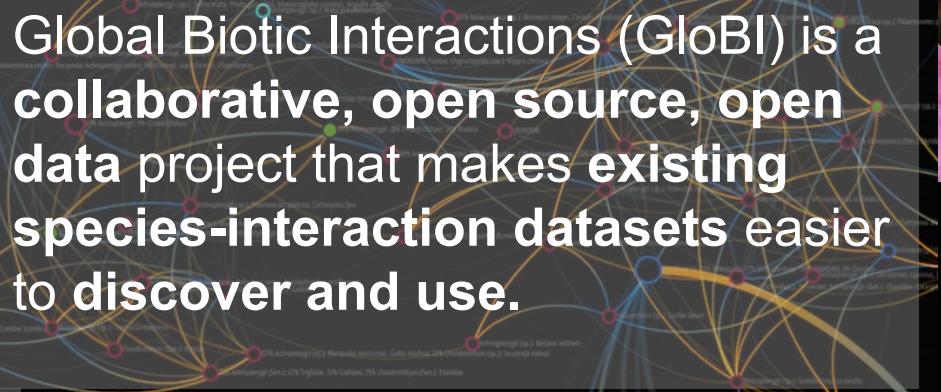
(c) edward_rooks, some rights reserved (CC BY-NC)
accessed at <http://www.inaturalist.org/observations/563486> on Feb 4, 2015



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

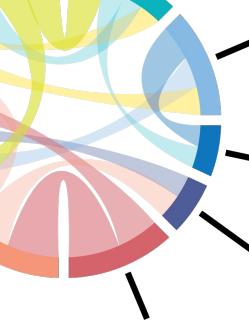


(c) Cheryl Harleston, some rights reserved (CC BY-NC-SA)
accessed at <http://www.inaturalist.org/observations/2020957> on Oct 13, 2015



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>



a brief history

2011 - Jorrit meets Jim at Texas A&M College Station for
Ecological Integration Symposium.

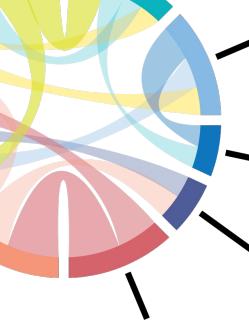
2013 - Encyclopedia of Life Rubenstein Fellowship /
GloBI / GoMexSI born

2014 - Jorrit, Jim and Chris publish GloBI paper

<many citations, workshops, conferences, integrations later>

2021 - Jorrit contributes to Dead Wood 2021

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>.



in a nutshell

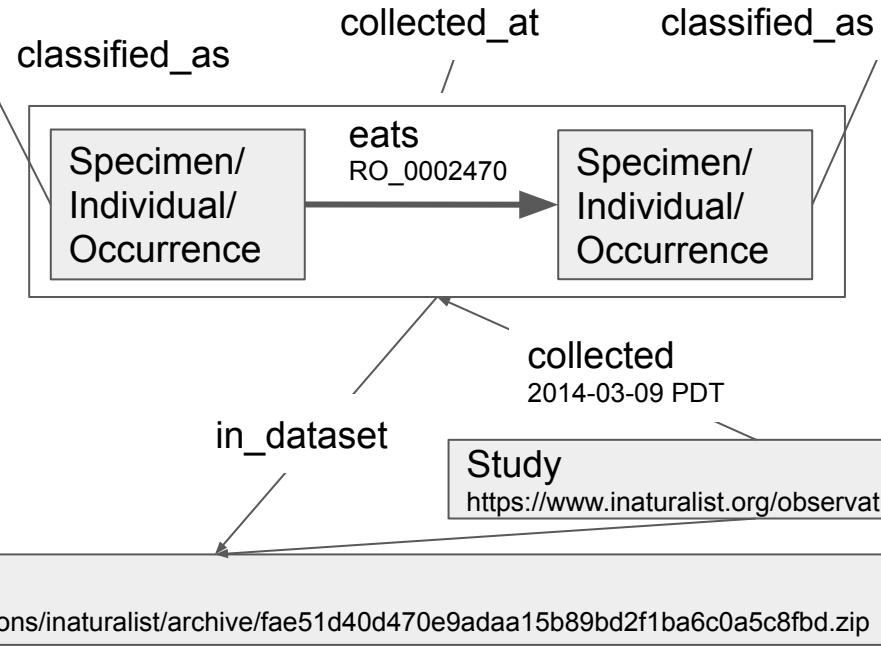
1. Existing openly accessible species interaction datasets in **any data format** are **registered** using
<https://globalbioticinteractions.org/contribute>
2. **GloBI** continually and automatically **indexes and links** most recent species interaction datasets.
3. **Users discover, access and improve** datasets via GloBI's many integrations, search pages, data archives or APIs.



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



Dataset
<https://github.com/globalbioticinteractions/inaturalist/archive/fae51d40d470e9adaa15b89bd2f1ba6c0a5c8fdb.zip>



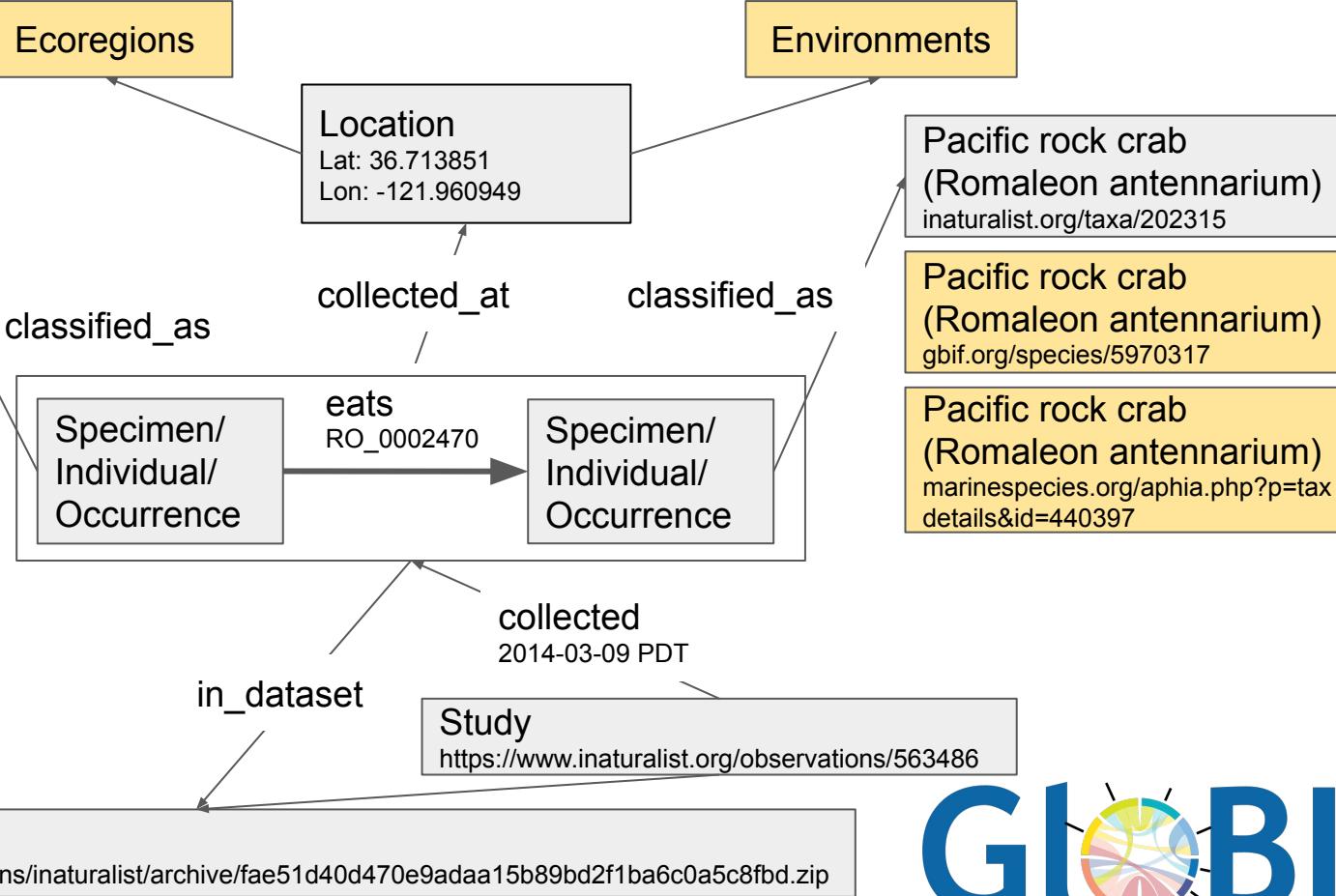
Simplified internal data model used by GloBI to integrate interaction data.



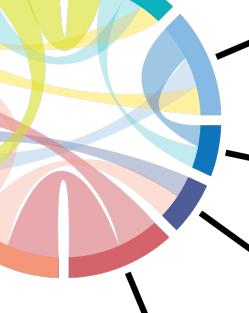
Southern Sea Otter
(*Enhydra lutris nereis*)
inaturalist.org/taxa/117520

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

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



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

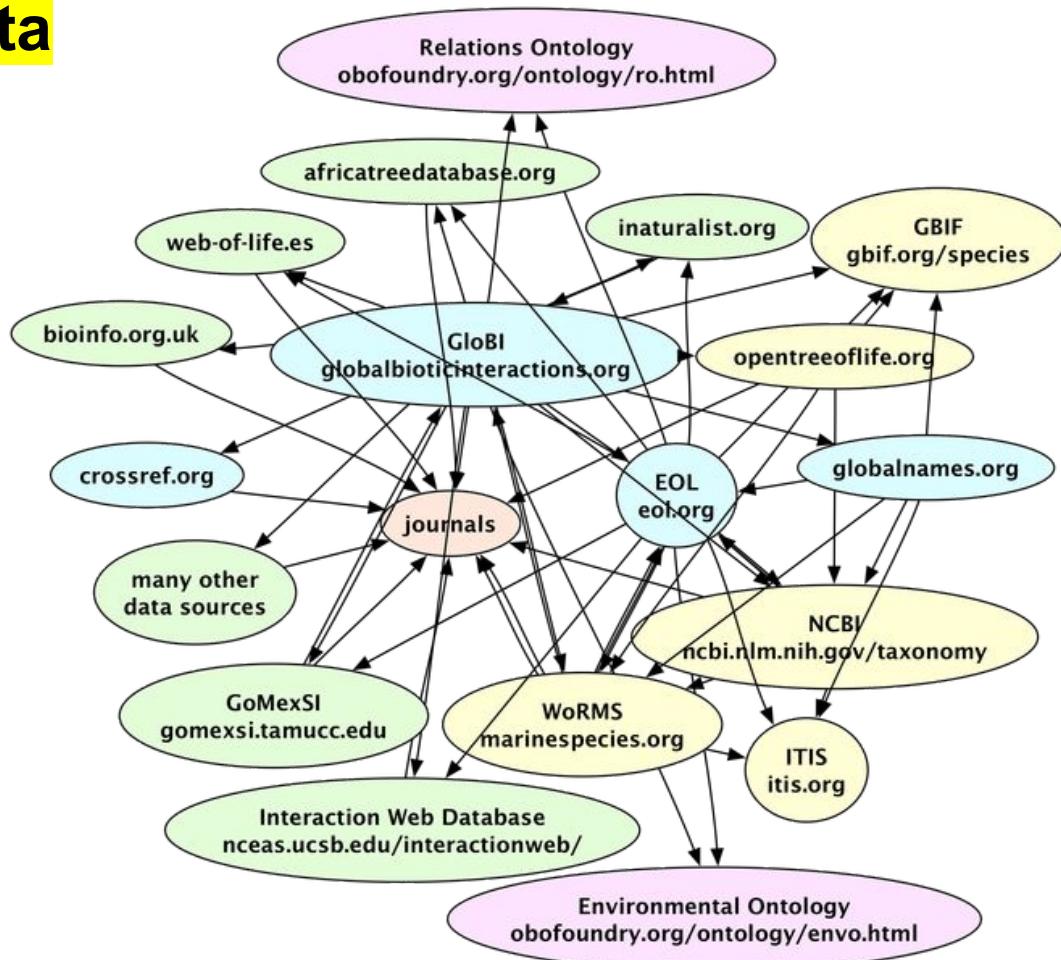


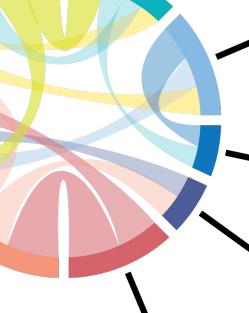
linking all the data

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.



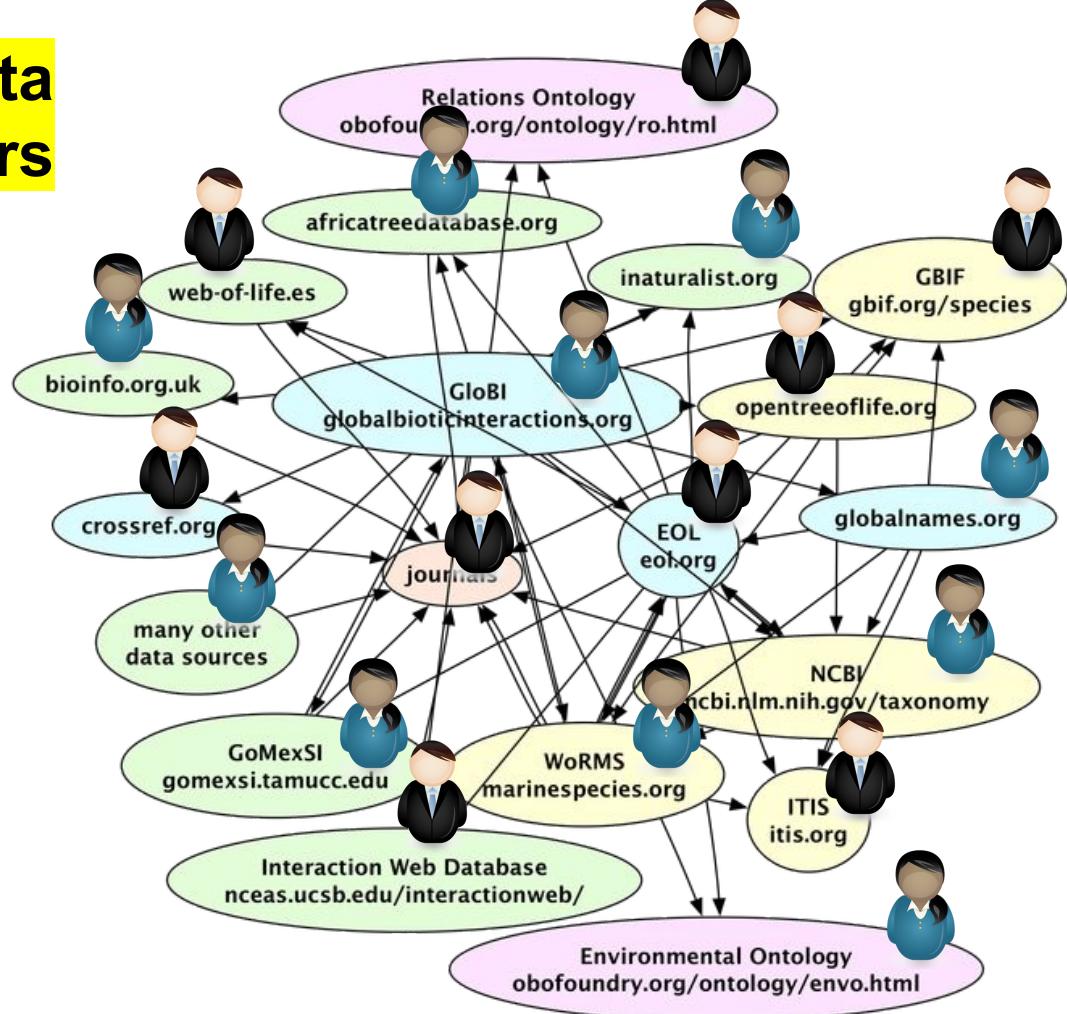


linking all the data and their curators

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.





it takes a village
to keep and share
a record

Data Mobilizers



Data Providers



Data Integration



Data Users



Ontology & Standards Communities

Taxonomists

Ecologists

OBO Foundry



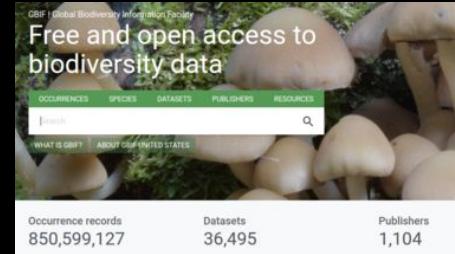
GloBI currently includes 268,317 references obtained from 310 data sources. In total, 2,852,596 interaction records were discovered, covering 181,570 taxa. A taxon map shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching 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 [terms / doi] claim that [some organism] interacts with [some other organism] ?

<http://globalbioticinteractions.org/references>

Accessed at 28 Sept 2017



<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>.

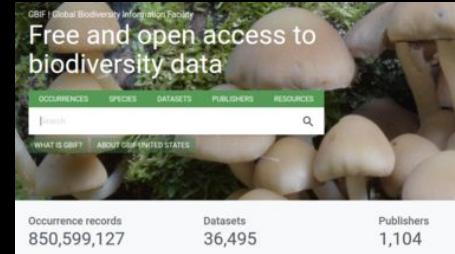
GloBI currently includes 293,203 references obtained from 326 data sources. In total, 3,379,426 interaction records were discovered, covering 233,557 taxa. A taxon map shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching 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 interacts with some other organism

<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>.

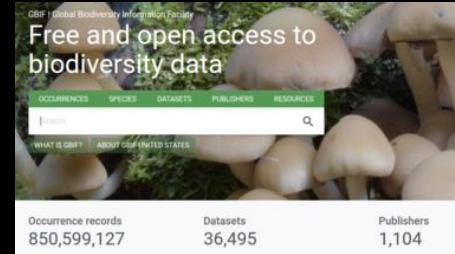
GloBI currently includes 293,203 references obtained from 326 data sources. In total, 3,379,426 interaction records were discovered, covering 233,557 taxa. A taxon map shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching 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 interacts with some other organism

<http://globalbioticinteractions.org/references>

Accessed at 9 Jan 2020



<https://gbif.org>

Accessed at 9 Jan 2020

4.4M records

0.2k datasets

~0.2M taxa

1.4B records

50.0k 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>.

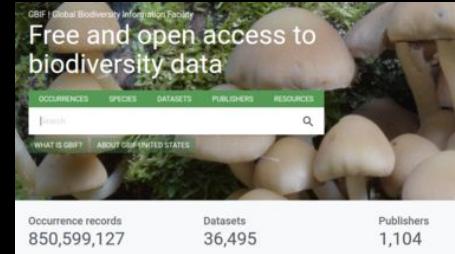
GloBI currently includes 293,203 references obtained from 326 data sources. In total, 3,379,426 interaction records were discovered, covering 233,557 taxa. A taxon map shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching 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 interacts with some other organism

<http://globalbioticinteractions.org/references>

Accessed at 11 April 2021



<https://gbif.org>

Accessed at 11 April 2021

6.8M records

0.3k datasets

~0.7M taxa

1.7B records

57.6k 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>.

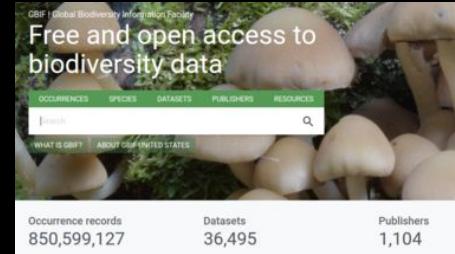
GloBI currently includes 293,203 references obtained from 326 data sources. In total, 3,379,426 interaction records were discovered, covering 233,557 taxa. A taxon map shows how these taxa relate to other projects (e.g. NCBI, WoRMS, EOL). Names that could not be linked by our automated taxon matching 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 interacts with some other organism ?

<http://globalbioticinteractions.org/references>

Accessed at 24 Oct 2021



<https://gbif.org>

Accessed at 24 Oct 2021

10.2M records

0.3k datasets

~0.7M taxa

1.9B records

62.8k datasets

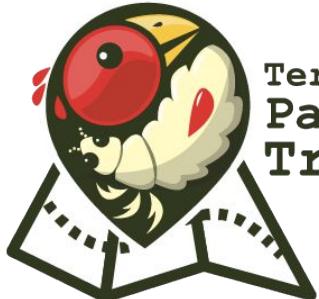
~1-2M species

Eltonian shortfall*: a lack of species-interaction records

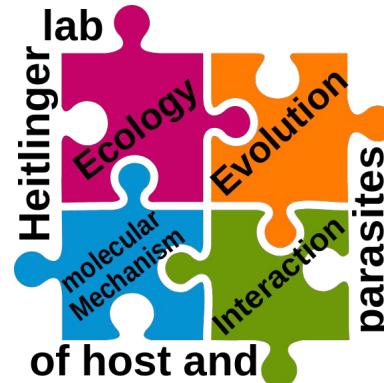
*Hortal, J. et al., 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46(1). Available at: <http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400>.



Encyclopedia of Life



partners



Article | Open Access | Published: 06 August 2019

Biological and environmental drivers of trophic ecology in marine fishes – a global perspective

B. Hayden M. L. D. Palomares, B. E. Smith & J. H. Poelen

Scientific Reports 9, Article number: 11415 (2019) | Cite this article

3077 Accesses | 41 Altmetric | Metrics



ANNUAL REVIEWS

JOURNALS
A-ZJOURNAL
INFOPRICING &
SUBSCRIPTIONS

SUBS

Home / Annual Review of Ecology, Evolution, and Systematics / Volume 46, 2015 / Hortal, pp 523-549

Em:

MENU ▾

nature

communications

Article | Open Access | Published: 29 November 2019

7-Deazaguanine modifications protect phage DNA from host restriction systems

Geoffrey Hutinet Witold Kot, Liang Cui, Roman Hillebrand, Seetharamsingh Balamkundu, Shanmugavel Gnanakalai, Ramesh Neelakandan, Alexander B. Carstens, Chuan Fa Lui, Denise Tremblay, Deborah Jacobs-Sera, Mandana Sasanfar, Yan-Jiun Lee, Peter Weigle, Sylvain Moineau, Graham F. Hatfull, Peter C. Dedon, Lars H. Hansen & Valérie de Crécy-Lagard

Nature Communications 10, Article number: 5442 (2019) | Cite this article

1162 Accesses | 25 Altmetric | Metrics

Research | Open Access |

Exploring the temporal variability of a food web using long-term biomonitoring data

Pierre Olivier Romain Frelat, Erik Bonsdorff, Susanne Kortsch, Ingrid Kröncke, Christian Möllmann, Hermann Neumann, Anne F. Sell, Marie C. Nordström

First published: 20 September 2019 | <https://doi.org/10.1111/ecog.04461>

ECOLOGY LETTERS

Review and Synthesis | Free Access |

Pyramids and cascades: a synthesis of food chain functioning and stability

Matthieu Barbier Michel Loreau

First published: 17 December 2018 | <https://doi.org/10.1111/ele.13196> | Citations: 5

PARASITE TRACKER

DOCUMENTING ARTHROPOD VERTEBRATE PARASITES



- **Project Title:** Collaborative Research: Digitization TCN: Digitizing collections to trace parasite-host associations and predict the spread of vector-borne disease
- **PIs:** Jen Zaspel & Stephen Cameron
- **Project Start Date:** September 1st 2019
- **Project Period:** 3 years
- **Participating Institutions:** 27
- **Co-PIs/Leads on Subs:** 34 (*12/17)
- **Participants:** 60+



Stephen Cameron, TPT Co-Lead



Jen Zaspel, TPT Co-Lead

PARASITE TRACKER

DOCUMENTING ARTHROPOD VERTEBRATE PARASITES



Neil Cobb, NAU



Rob Guralnick, UF



Katja Seltmann, UCSB Julie Allen, UNR



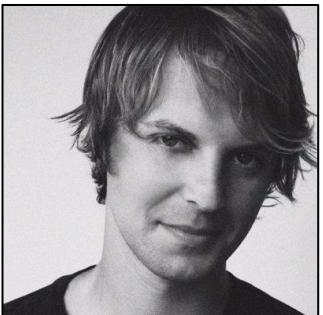
Julie Allen, UNR



Jorrit Poelen, GloBI



Mark Smith, MS LLC



André Poremski, Fg Anna Monfils, BLUE



Alyssa Caywood, MPM



Chris Tyrell,
MPM



Julia Colby,
MPM



Barb Thiers, NYBG



Matt Yoder,
TaxonWorks

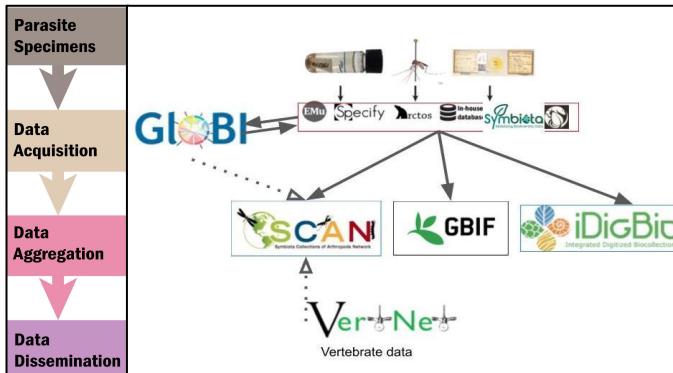
PARASITE TRACKER

DOCUMENTING ARTHROPOD VERTEBRATE PARASIT



Terrestrial Parasite Tracker TCN

- Transcribe and georeference label data from **1.2+** million arthropod parasite specimens from 22 collections across North America (U.S. and territories)
- Document 500,000+ parasite-host associations via GloBI



PARASITE TRACKER

DOCUMENTING ARTHROPOD VERTEBRATE PARASITES



Data Integration

Global Biotic Interactions (GloBI) is a data integration tool that indexes existing species interaction datasets, literature, and specimen records in collections.

<https://www.globalbioticinteractions.org/>

Specimen data transcribed for the TPT project will generate 500,000 new parasite-host association records in GloBI.



about blog browse contribute data search references status 日本語 Español

Example query: What do sea otters (*Enhydra lutris*) eat? or What do honey bees (*Apis*) pollinate?

What kind of Siphonaptera do organisms interacts with according to any study or source

organisms
Interacts with... plenty of things!



Elder
(Somateria)

...

has parasite



fleas
(Siphonaptera)

...

Supported by:

<http://invertebrates.si.edu/parasites.htm>. Accessed at <cleaned_up.tsv> on 16 Nov 2019. show Provider:
<http://invertebrates.si.edu/parasites.htm>. Accessed at <cleaned_up.tsv> on 16 Nov 2019.

Benesh, D. P., Lafferty, K. D. and Kuris, A. (2017). A life cycle database for parasitic acanthocephalans, cestodes, and nematodes. *Ecology*, 98: 882. doi:10.1002/ecy.1680 [link](#) show Provider: Sarah E Miller. 9/19/2017. Species associations manually extracted from Benesh, D. P., Lafferty, K. D. and Kuris, A. (2017). A life cycle database for parasitic acanthocephalans, cestodes, and nematodes. *Ecology*, 98: 882. doi:10.1002/ecy.1680. Accessed at <<https://github.com/millerse/Benesh-et-al-2017/archive/9dd51ac0c04ab7b06761d30032d2b93369855fd.zip>> on 16 Nov 2019.

<http://arctos.database.museum/guid/MSB:Para:16981> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:2861> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:3133> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:3418> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:872> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:5340> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:16984> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:3463> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:3177> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32

<http://arctos.database.museum/guid/MSB:Para:5341> [link](#) show Provider: MSB Parasite Collection (Arctos) - Version 32.32



<https://globalbioticinteractions.org/parasitetracker>

Indexing and Reviewing Research Datasets

TPT Collections Status

Click on badges to browse/download indexed records or inspect automated reviews.

[edit collection list](#)

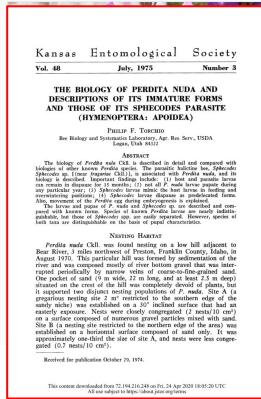


status	Institution/collection	platform	contact
	BPBM / Bishop Museum / J. Linsley Gressitt Center for Research in Entomology	Specity	Jim Boone, Neal Evenhuis
	BYU / Brigham Young University Arthropod Museum BYUC / Monte L. Bean Museum of Life Science Museum Insects and Arachnids	SymbSCAN	Michael Whiting
	CAS / California Academy of Sciences / Entomology Collection	Institutional	Chris Grinter, Michelle Trautwein
	CUAC / Clemson University / Clemson University Arthropod	Parasitetracker	David Seltmann

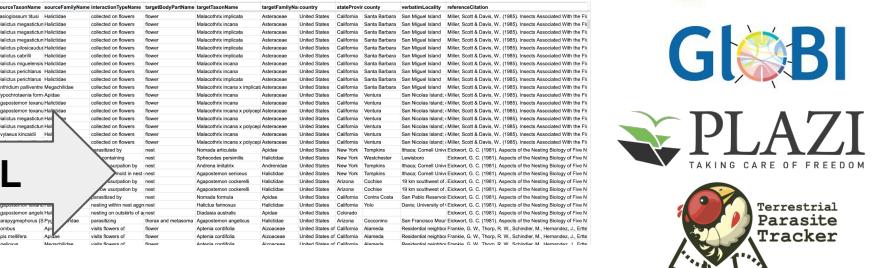
The screenshot displays the GlobI interface with several key components:

- Review Notes:** A callout box highlights a review note for the 'globi' collection, mentioning 14024 interactions, 48 notes, and 14024 info(s).
- Indexed Interactions Tables:** A callout box shows a table structure with a grid of icons.
- Indexed Names Tables:** A callout box shows a table structure with a grid of icons.
- Browse Explore Discover:** A large orange box containing the text "Browse Explore Discover".
- Configuration Translation Tables Data Location:** A large orange box containing the text "Configuration Translation Tables Data Location".
- Discussion Questions Notes:** A large orange box containing the text "Discussion Questions Notes".
- GlobI Status Badges:** A vertical stack of status badges for review, GlobI, config, issues (3 open), and names.
- GitHub Integration:** A GitHub repository card for 'globalbioticinteractions / cas-ent' is shown, including code, issues (3), pull requests, and actions.

Creating Research Datasets: Directed Research in Data Science for Undergraduates



>>>>> GOAL



Understandable by people and computers

[ParasiteTracker / tick-interaction-database](#)

Code Issues Pull requests Actions Projects Wiki Security Insights

master · 1 branch · 0 tags

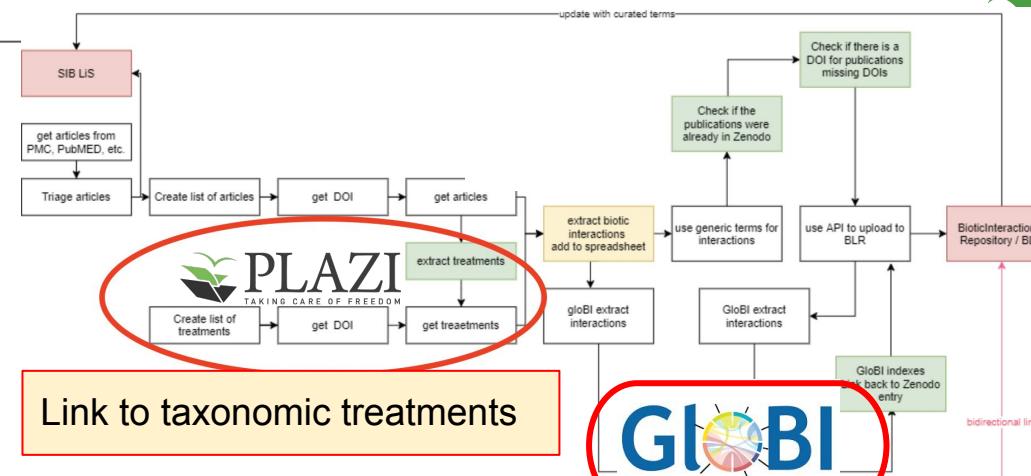
Get file Add file Code

seitmann Merge branch 'master' of https://github.com/ParasiteTracker/ti... 89944e1 18 days ago 13 commits

- .travis.yml modified README and added .travis 18 days ago
- README.md Merge branch 'master' of https://github.com/ParasiteTracker/ti... 18 days ago
- globi.json fix globi.json for Tick Interaction Database 22 days ago
- interaction_types_mapping.csv initial commit copied from BID 4 months ago
- interactions.tsv tick interactions from the literature 25 days ago

README.md

Contributors 6



Bee Interaction & Tick Interaction Databases

Data review with GLOBI

Katja Seltmann,
seltmann@ucsb.edu



Management of biotic interaction data in the Encyclopedia of Life

Katja Schulz (SchulzK@si.edu)¹, Jennifer Hammock (HammockJ@si.edu)¹, Jorrit Poelen (jhpoelen@xs4all.nl)², Eli Agbayani (eagbayani@eol.org)¹

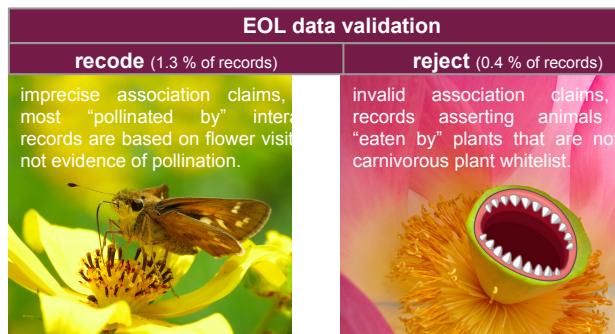
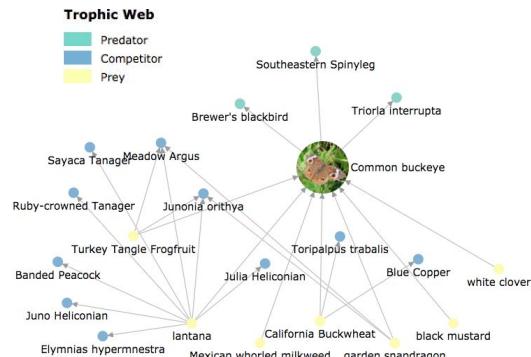
¹Encyclopedia of Life, Smithsonian National Museum of Natural History, Washington, DC, USA

²Global Biotic Interactions, Oakland, California, USA

Species interactions are an important component of TraitBank, the structured data resource developed by the Encyclopedia of Life (EOL, eol.org/traitbank). EOL provides *human- and machine-accessible interfaces* to query and download the data, and *interactive food web visualizations* are available on EOL species pages.

The major supplier of EOL interaction data is Global Biotic Interactions (GloBI, globalbioticinteractions.org), an open-source infrastructure for the efficient integration & sharing of interaction data from diverse sources.

The quality of GloBI data is heterogeneous. Data from text mining or citizen science projects may have undergone little or no curation, and records derived from the peer-reviewed scientific literature may be subject to errors in transcription or interpretation. Many data problems are caused by incorrect taxon mappings due to homonymy or invalid taxon hierarchies. EOL therefore applies a series of validation rules to ensure interaction data are fit for EOL use.



639 data sources • 4,891,426 interaction records • 331,512 taxa
scientific literature | databases | natural history collections | citizen science projects | text mining



interaction records refuted by EOL

DwC-A with association extension



To alert other GloBI users of potential data problems, rejected association claims are fed back to GloBI in the form of "*refuted association records*". Best practices for the expression and leveraging of refuted associations are still very much under development. Anybody interested in these efforts can participate in the discussion at github.com/globalbioticinteractions/refuted-interaction-data

Liberating Biodiversity Data From COVID-19 Lockdown:

Toward a knowledge hub for mammal host-virus information

Presenter: Nate Upham,
Arizona State University

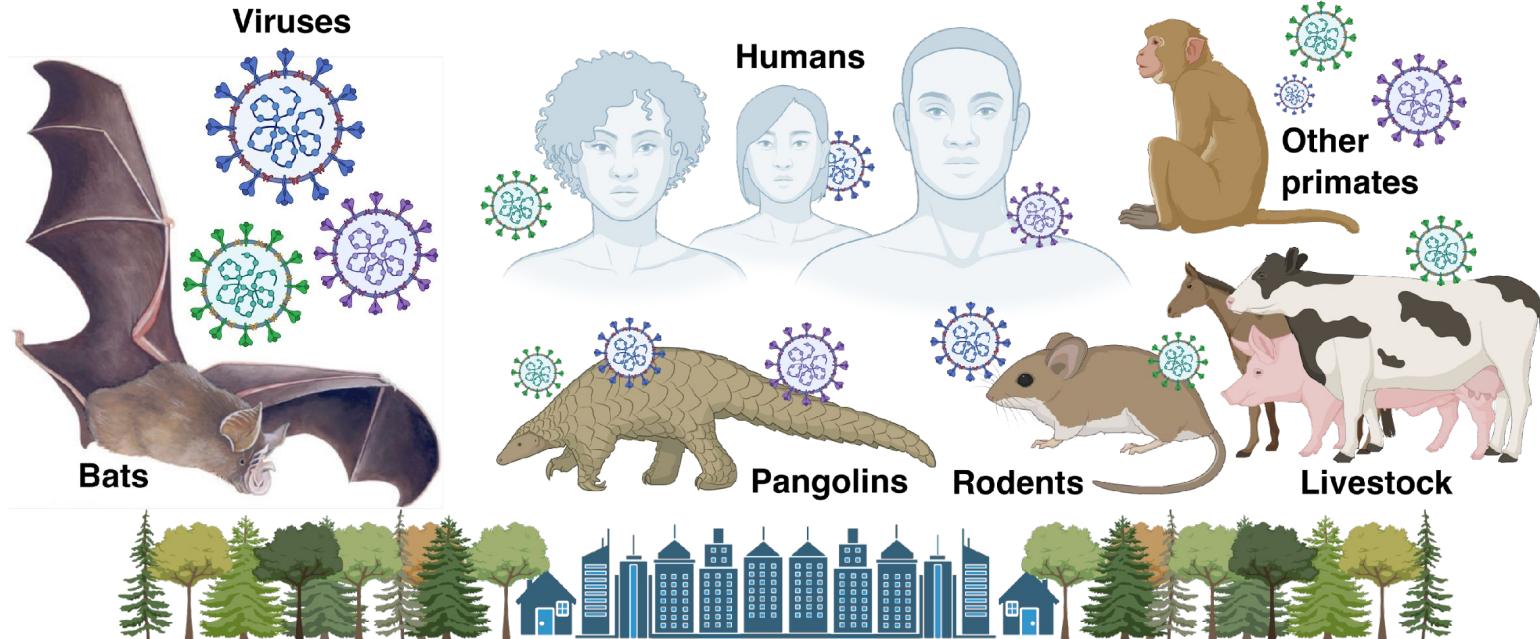
Contributors: Donat Agosti, Jorrit Poelen, Lyubomir Penev, Deborah Paul, DeeAnn M. Reeder, Nancy B. Simmons, Gabor Csorba, Quentin Groom, Mariya Dimitrova, Joseph T Miller



CETAF-DISSCO COVID-19 TASKFORCE



Viral spillover risk from mammals to humans

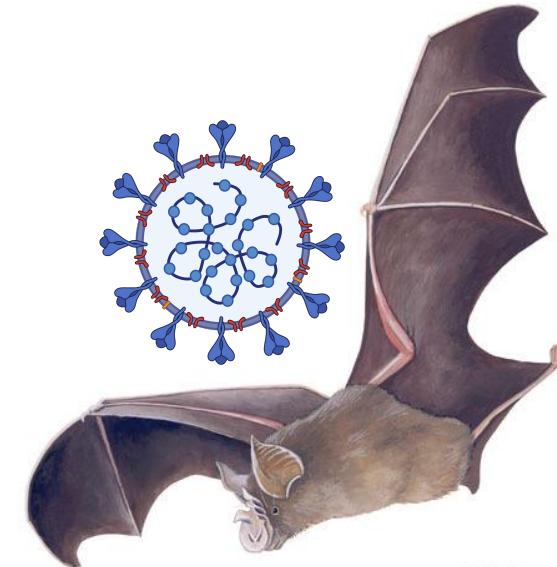


April 2020... Goal: liberate host-virus data

- **Novel coronavirus:** SARS-CoV-2
- **Likely host reservoirs:** mammals > bats > horseshoe bats (*Rhinolophus* sp.)



Nancy Simmons
AMNH Mammal Curator
Global bat expert
batnames.org



Analog articles to digitally extracted data

Get publications:

- PDF (library access)

Receive publications

- PD
 - Born digital
 - Scanned
- XML
 - JATS
 - TaxPub/JATS



Import

- manual
- automated

Enhance

- link
- semantics



SynoSpecies



Create open FAIR data

- TreatmentBank
- BLR / Zenodo

Access / Metrics

- User Interface
- API
- Dashboards etc

Research
questions



BLR: [Biodiversity Literature Repository](#) / [Zenodo](#) / [CoViHo](#)

FAIR: [Findable, Accessible, Interoperable, Reusable](#)

GBIF: [Global Biodiversity Information Facility](#)

JATS: [Journal Archival Tag Suite](#)

LINDAS: [Linked Data Service Switzerland](#)

NCBI: [National Center for Biotechnology Information](#)

PDF: Portable Document Format

OpenBiodiv: [Open Biodiversity Knowledge Management System](#)

SIBILS: [Swiss Institute of Bioinformatics Library System](#)

Synospecies: [Taxonomic Name Resolution Service](#)

TaxPub: [Taxonomic extension of JATS](#)

Wikidata: [Wikidata](#)

XML: Extended Markup Language (Documents)

Semantic Publishing Enables Text Mining of Biotic Interactions

Mariya Dimitrova, Jorrit Poelen, Georgi Zhelezov, Teodor Georgiev, Donat Agosti, Lyubomir Penev

Presenter: Mariya Dimitrova



PENSOFT[®]



Old articles



Journal Articles & Figures

Legacy literature



Text and data mining



Manuscripts



Peer review, semantic tagging,
publication, distribution

zenodo



*Research
questions*



- Publish articles as digital, **semantically linked** from the start
- Semantic publishing:
 - Data is exposed in a structured way (XML, JSON) so that others can use it
 - Digitally accessible in a translatable vocabulary

Extracted data to indexed ecological networks



From Zenodo:

- Via Plazi, Pensoft

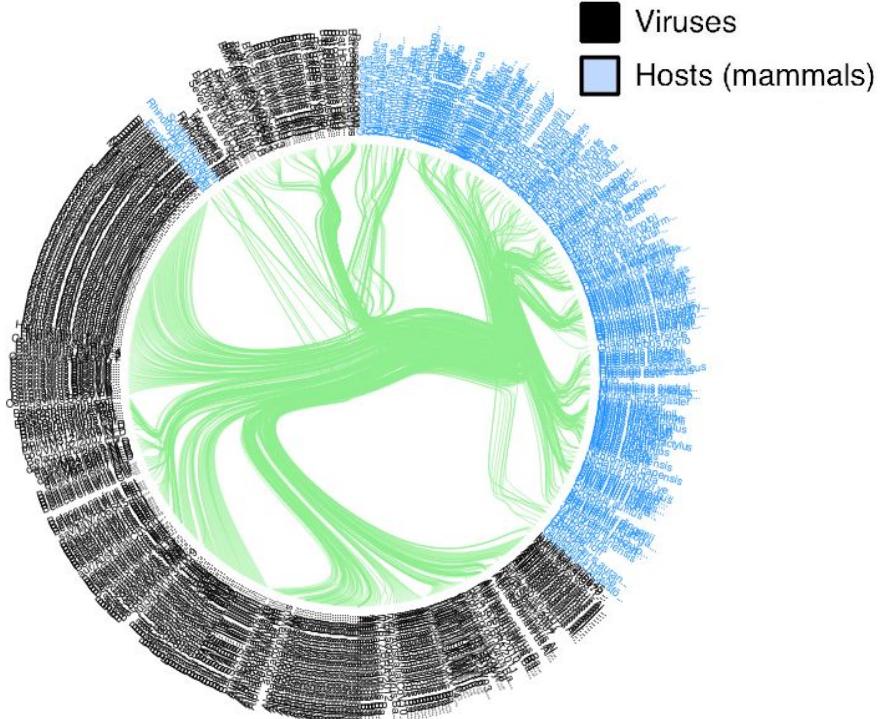
Newly indexed data:

- BatBase/BatPlants
- DBatVir
- DRodVir
- NCBI Virus



Existing datasets:

- PHI-Base, Virus Host DB



Outcomes:



Global Biotic Interactions

zenodo

Search

Upload Communities

jhpoele@xs4all.nl

October 6, 2020

Dataset Open Access

CETAF-DiSSCo/COVID19-TAF biodiversity-related knowledge hub working group: indexed biotic interactions and review summary

Poelen, Jorrit; Upham, Nathan; Agosti, Donat; Ruschel, Tatiana; Guidoti, Marcus; Reeder, DeeAnn; Simmons, Nancy; Penev, Lyubomir; Dimitrova, Mariya; Csorba, Gabor; Groom, Quentin; Willoughby, Anna

This data publication originated as part of developing a biodiversity-related knowledge hub on COVID-19 via COVID19-TAF - Communities Taking Action (<https://cetaf.org/covid19-taf-communities-taking-action>), a community-rooted initiative raised jointly by the Consortium of European Taxonomic Facilities (CETAF, <https://cetaf.org>) and Distributed Systems of Scientific Collections (DiSSCo, <https://www.dissco.eu/>).

This archive contains the biodiversity datasets of interest identified in period 14 April-6 October 2020 through COVID19-TAF activities and subsequently indexed by Global Biotic Interactions (GloBI, <https://globalbioticinteractions.org>). GloBI provides open access to finding species interaction data (e.g., predator-prey, pollinator-plant, virus-host, parasite-host) by combining existing open datasets using open source software.

These identified datasets (see references and reviews below) add to a growing collection of open species interaction datasets already indexed by GloBI. So, this data publication only includes a small subset of indexed datasets and include

Edit

New version

Communities

Coronavirus Disease Research Community - COVID-19

Zenodo

854 views

314 downloads

[See more details...](#)

Poelen, Jorrit, Upham, Nathan, Agosti, Donat, Ruschel, Tatiana, Guidoti, Marcus, Reeder, DeeAnn, ... Willoughby, Anna. (2020). CETAF-DiSSCo/COVID19-TAF biodiversity-related knowledge hub working group: indexed biotic interactions and review summary (Version 0.3) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4068958>

Acknowledgments

an incomplete list in no particular order

These projects would not have been 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, Dmitri Mozzherin, Malcolm Storey, Michael Elliott, José Fortes, Quentin Groom, Nathan Upham, Deb Paul, Mariya Dimitrova, Lyubomir Penev, Donat Agosti, Alex Ioannidis, Marcus Guidoti, Jen Zaspel, Carl Boettiger ... and thousands of others that have collected and shared biodiversity data and related tools.

Thank you!

Jorrit Poelen @ <https://jhpoelen.nl>

Works presented are funded in part by grants OAC 1839201, DBI 1901932, DBI 1901926 from the National Science Foundation, the Encyclopedia of Life Rubenstein Fellows Program (CRDF EOL-33066-13/F33066, 2013), EOL David M. Rubenstein Grant (FOCX-14-60988-1, 2014), and the Smithsonian Institution (SI) (T15CC10297-002, 2016).

Extra

Open discussions and reviews using GitHub issues

The screenshot shows a GitHub repository interface. On the left, there's a sidebar with a list of issues, some of which are highlighted with orange boxes and arrows pointing to specific comments in the main area. The main navigation bar includes 'Code', 'Issues 62' (which is highlighted), 'Pull requests 0', 'Projects 0', 'Wiki', 'Settings', and 'Insights'. Below the navigation, the repository title is 'A life cycle database for parasitic acanthocephalans, cestodes, and nematodes by Benesh et al. #305'. A green button labeled 'Open' indicates the issue is open. The issue itself is from user 'jhpoelen' on Aug 14, with 16 comments. A comment by 'jhpoelen' is expanded, showing a reply from '@derele'. The comment text reads:

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

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

As suggested by **@derele** -

[...]

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

[...]



Emanuel Heitlinger
derele

<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.

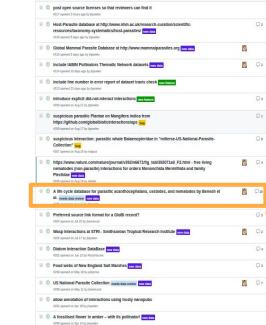


millerse commented 9 days ago



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

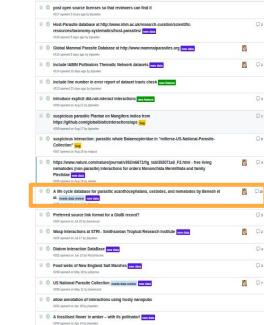
Collaborator



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

1. post open source formats so that reviewers can find it
2. add Provenance information to GitHub issues and comments, referencing external sources like GitHub pull requests
3. Create GitHub Issues Thread Network accounts
4. Create GitHub Issues Thread Network accounts
5. Add GitHub Issues Thread Network accounts
6. Add GitHub Issues Thread Network accounts
7. Add GitHub Issues Thread Network accounts
8. Add GitHub Issues Thread Network accounts
9. Add GitHub Issues Thread Network accounts
10. Add GitHub Issues Thread Network accounts
11. Add GitHub Issues Thread Network accounts
12. Add GitHub Issues Thread Network accounts
13. Add GitHub Issues Thread Network accounts
14. Add GitHub Issues Thread Network accounts
15. Add GitHub Issues Thread Network accounts
16. Add GitHub Issues Thread Network accounts
17. Add GitHub Issues Thread Network accounts
18. Add GitHub Issues Thread Network accounts
19. Add GitHub Issues Thread Network accounts
20. Add GitHub Issues Thread Network accounts
21. Add GitHub Issues Thread Network accounts
22. Add GitHub Issues Thread Network accounts
23. Add GitHub Issues Thread Network accounts
24. Add GitHub Issues Thread Network accounts
25. Add GitHub Issues Thread Network accounts
26. Add GitHub Issues Thread Network accounts
27. Add GitHub Issues Thread Network accounts
28. Add GitHub Issues Thread Network accounts
29. Add GitHub Issues Thread Network accounts
30. Add GitHub Issues Thread Network accounts
31. Add GitHub Issues Thread Network accounts
32. Add GitHub Issues Thread Network accounts
33. Add GitHub Issues Thread Network accounts
34. Add GitHub Issues Thread Network accounts
35. Add GitHub Issues Thread Network accounts
36. Add GitHub Issues Thread Network accounts
37. Add GitHub Issues Thread Network accounts
38. Add GitHub Issues Thread Network accounts
39. Add GitHub Issues Thread Network accounts
40. Add GitHub Issues Thread Network accounts
41. Add GitHub Issues Thread Network accounts
42. Add GitHub Issues Thread Network accounts
43. Add GitHub Issues Thread Network accounts
44. Add GitHub Issues Thread Network accounts
45. Add GitHub Issues Thread Network accounts
46. Add GitHub Issues Thread Network accounts
47. Add GitHub Issues Thread Network accounts
48. Add GitHub Issues Thread Network accounts
49. Add GitHub Issues Thread Network accounts
50. Add GitHub Issues Thread Network accounts
51. Add GitHub Issues Thread Network accounts
52. Add GitHub Issues Thread Network accounts
53. Add GitHub Issues Thread Network accounts
54. Add GitHub Issues Thread Network accounts
55. Add GitHub Issues Thread Network accounts
56. Add GitHub Issues Thread Network accounts
57. Add GitHub Issues Thread Network accounts
58. Add GitHub Issues Thread Network accounts
59. Add GitHub Issues Thread Network accounts
60. Add GitHub Issues Thread Network accounts
61. Add GitHub Issues Thread Network accounts
62. Add GitHub Issues Thread Network accounts
63. Add GitHub Issues Thread Network accounts
64. Add GitHub Issues Thread Network accounts
65. Add GitHub Issues Thread Network accounts
66. Add GitHub Issues Thread Network accounts
67. Add GitHub Issues Thread Network accounts
68. Add GitHub Issues Thread Network accounts
69. Add GitHub Issues Thread Network accounts
70. Add GitHub Issues Thread Network accounts
71. Add GitHub Issues Thread Network accounts
72. Add GitHub Issues Thread Network accounts
73. Add GitHub Issues Thread Network accounts
74. Add GitHub Issues Thread Network accounts
75. Add GitHub Issues Thread Network accounts
76. Add GitHub Issues Thread Network accounts
77. Add GitHub Issues Thread Network accounts
78. Add GitHub Issues Thread Network accounts
79. Add GitHub Issues Thread Network accounts
80. Add GitHub Issues Thread Network accounts
81. Add GitHub Issues Thread Network accounts
82. Add GitHub Issues Thread Network accounts
83. Add GitHub Issues Thread Network accounts
84. Add GitHub Issues Thread Network accounts
85. Add GitHub Issues Thread Network accounts
86. Add GitHub Issues Thread Network accounts
87. Add GitHub Issues Thread Network accounts
88. Add GitHub Issues Thread Network accounts
89. Add GitHub Issues Thread Network accounts
90. Add GitHub Issues Thread Network accounts
91. Add GitHub Issues Thread Network accounts
92. Add GitHub Issues Thread Network accounts
93. Add GitHub Issues Thread Network accounts
94. Add GitHub Issues Thread Network accounts
95. Add GitHub Issues Thread Network accounts
96. Add GitHub Issues Thread Network accounts
97. Add GitHub Issues Thread Network accounts
98. Add GitHub Issues Thread Network accounts
99. Add GitHub Issues Thread Network accounts
100. Add GitHub Issues Thread Network accounts



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

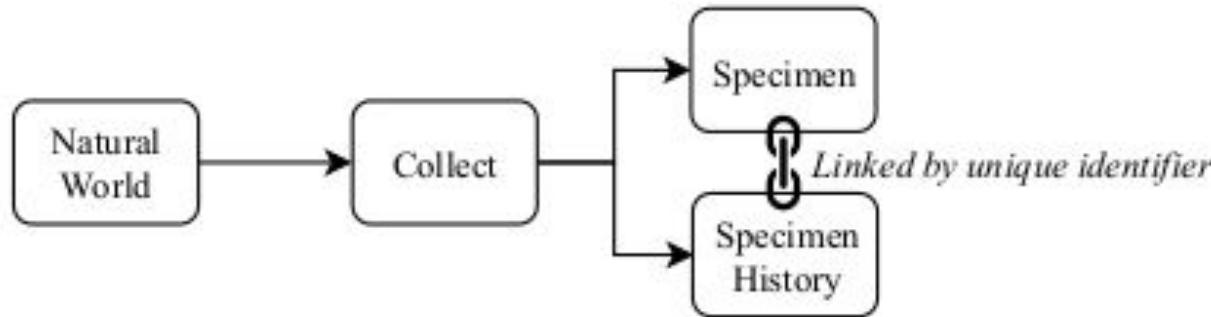


Okay everyone. I have added the endoparasite link and the body terms to the sheet.

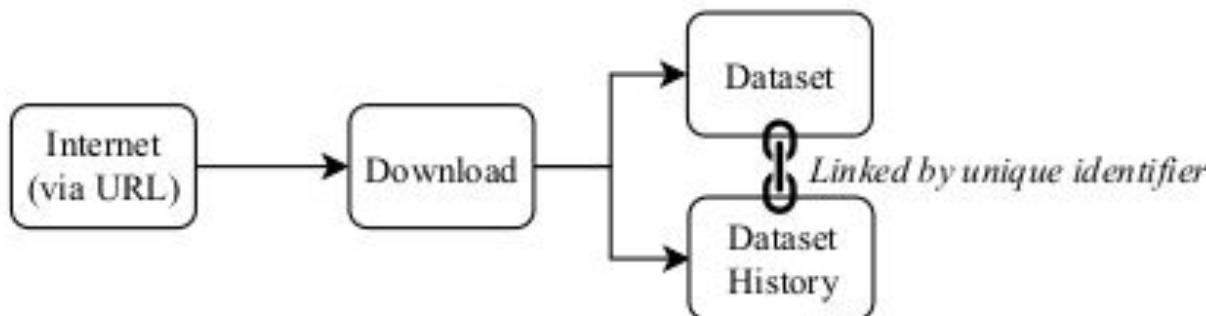
@dbenesh82 To your comment about the extra numbers: You had multiple body locations in a single record (i.e. gut, body cavity, etc.). I took each of those terms and made a single full record for each term. That is where the records come from.



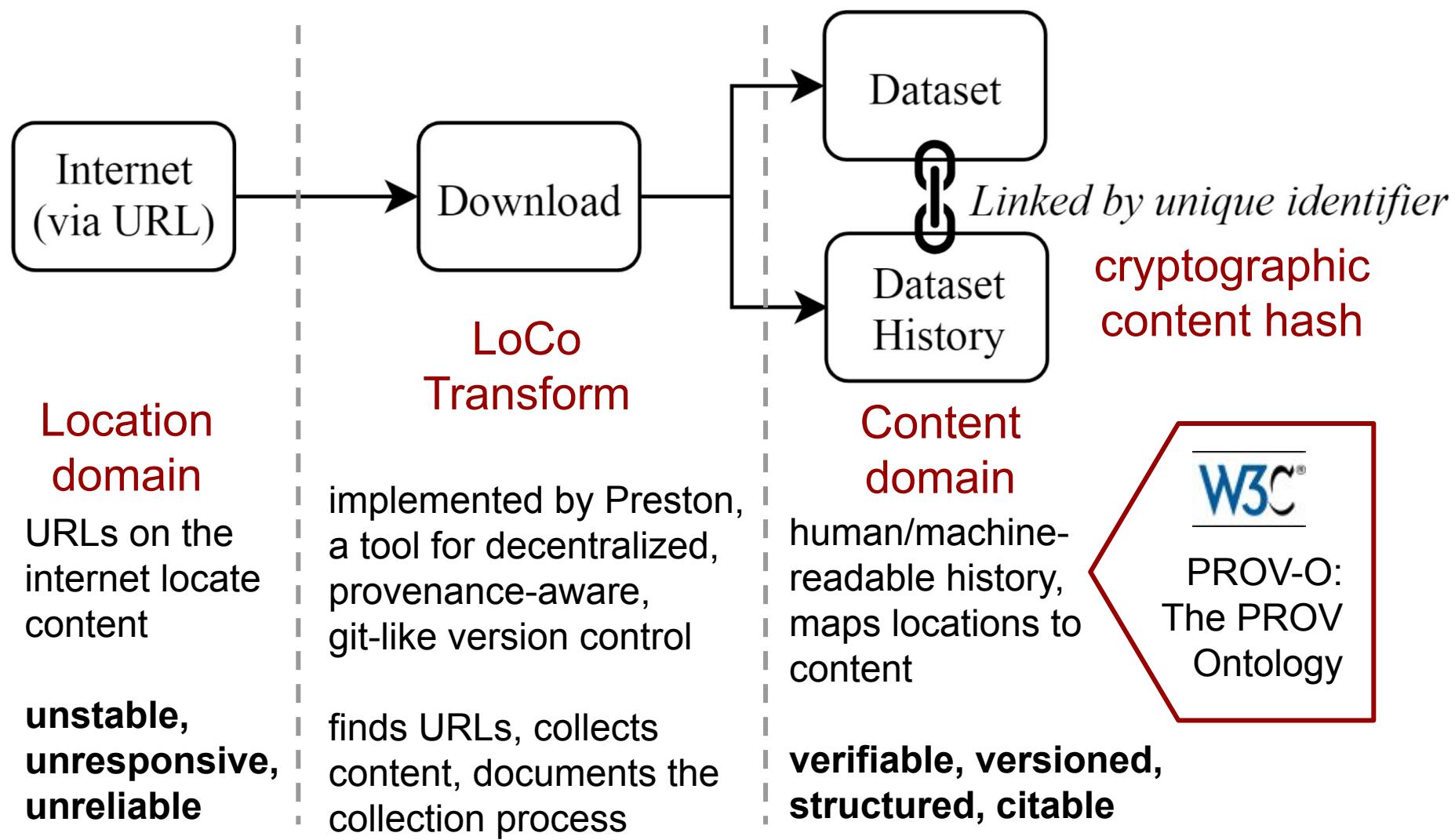
-□- Total URLs
··◆·· Total Contents

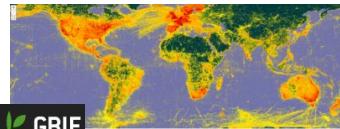
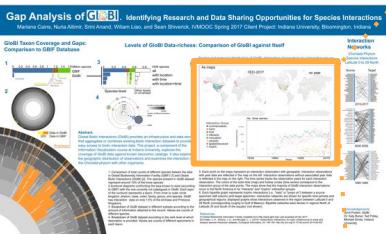


(a) Physical specimen collection



(b) Digital data collection

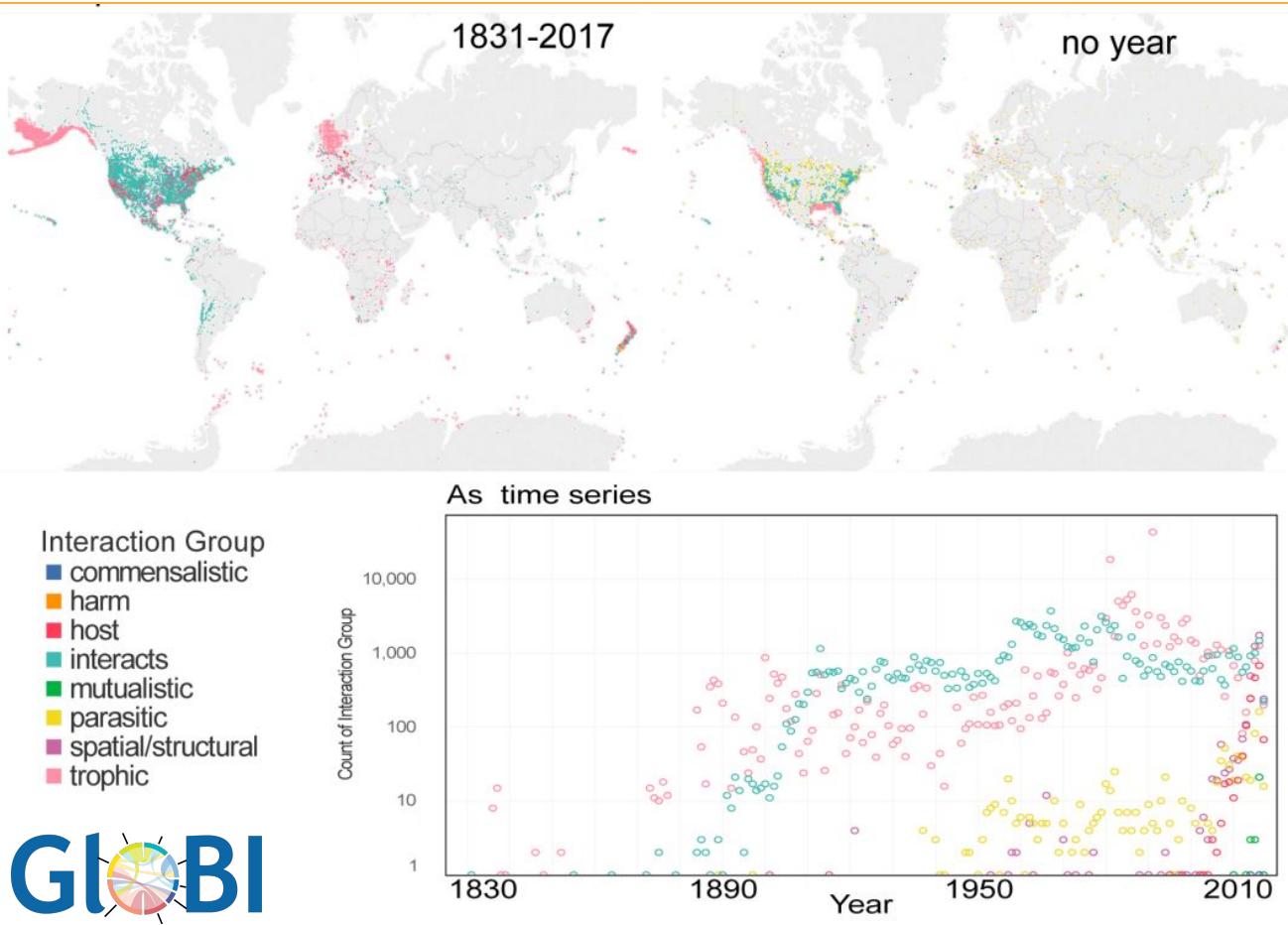


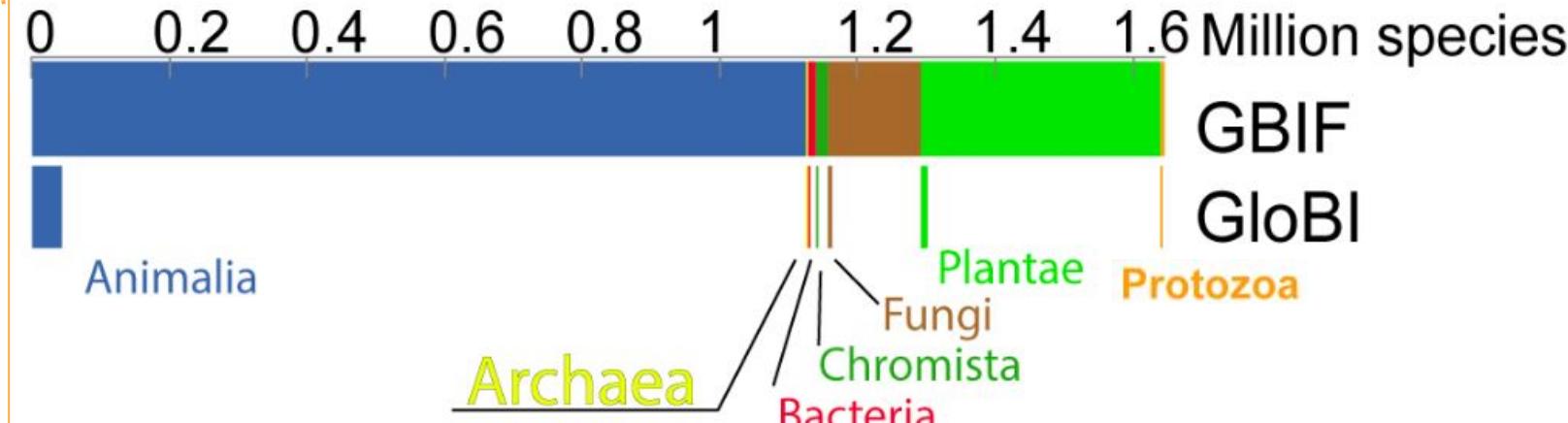
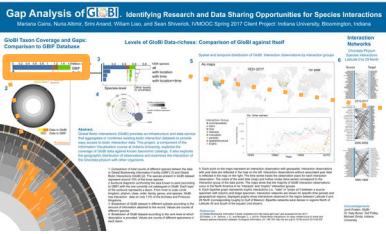


geospatial
and
temporal
Eltonian
shortfall



- Interaction Group
- commensalistic
 - harm
 - host
 - interacts
 - mutualistic
 - parasitic
 - spatial/structural
 - trophic





taxonomic Eltonian shortfall