On keeping track of parasite-host records

Jorrit Poelen, GloBI & Ronin Institute

@ Terrestrial Parasite Tracker workshop

23-25 Feb 2020 Field Museum, Chicago, IL

How can GloBI help to share your parasite-host records?

How can ~~GloBI~~ you help to share your parasite-host records?

How can ~~GloBI~~ ~~you~~ **we** help **each other** to share ~~your~~ **our** parasite-host records?

How can **we** help **each other** to share **our** parasite-host records?

topics

> Vision and Context
> Integration Profiles
> Reviews and Indexes
> Interaction Terms and Translation Tables

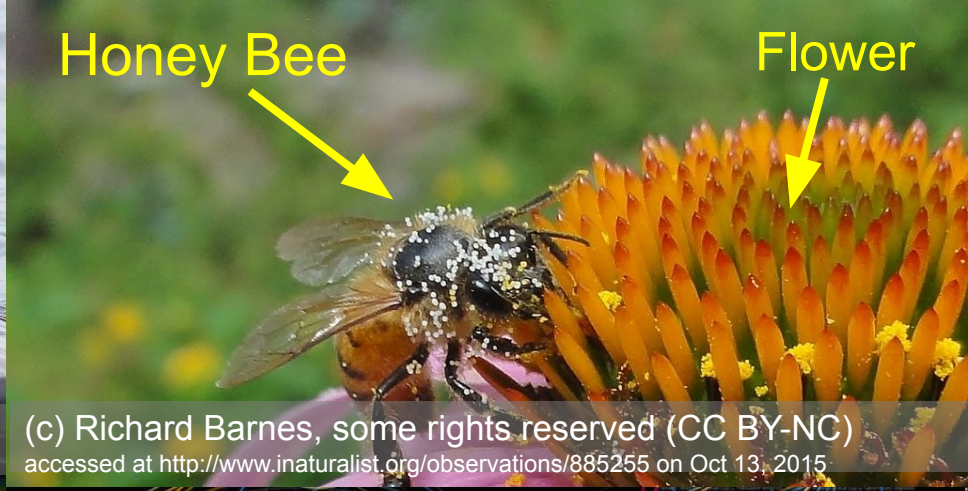
topics

> Vision and Context

> Integration Profiles

> Reviews and Indexes

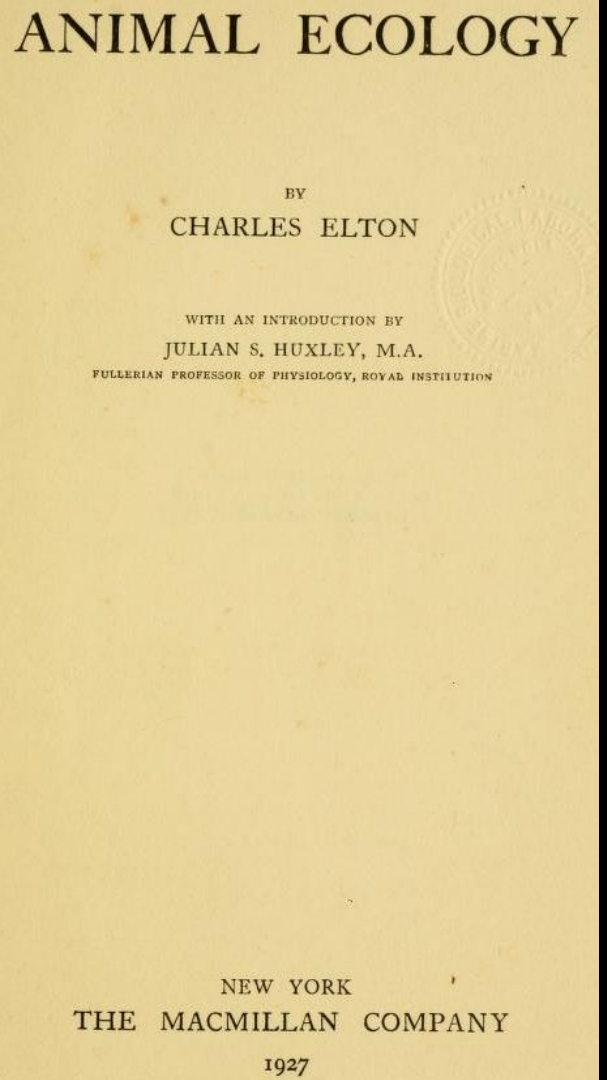
> Interaction Terms and Translation Tables



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


ANIMAL ECOLOGY

BY

CHARLES ELTON

WITH AN INTRODUCTION BY

JULIAN S. HUXLEY, M.A.

FULLERIAN PROFESSOR OF PHYSIOLOGY, ROYAL INSTITUTION

NEW YORK

THE MACMILLAN COMPANY

1927

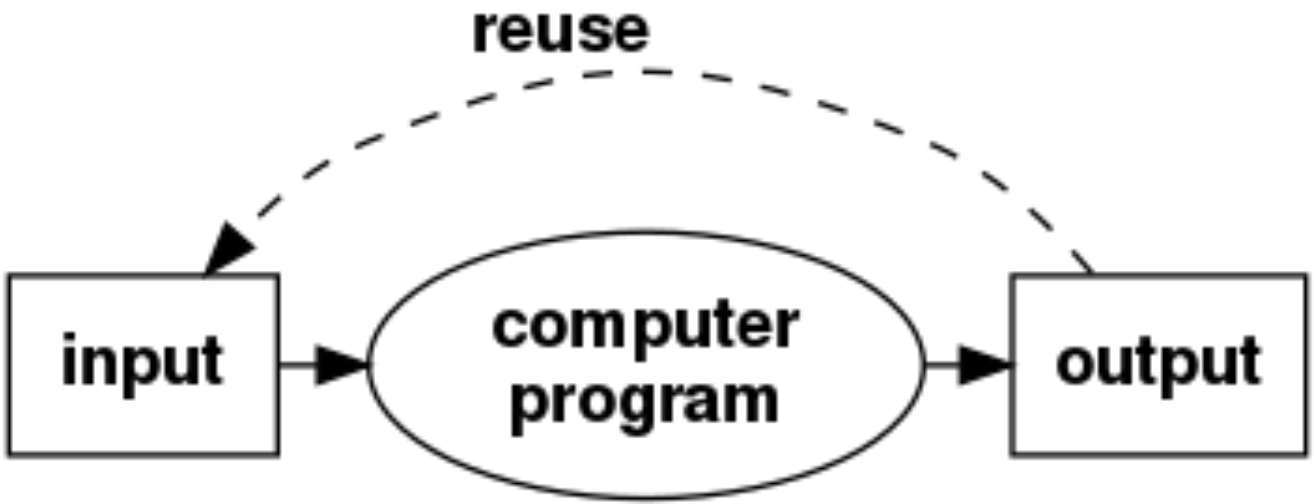
“The advantage, and at the same time the difficulty, of ecological work is that it attempts to provide conceptions which can link up into some complete scheme the colossal store of facts about natural history which has accumulated up to date in this rather haphazard manner. [...] Until more organised 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*.

Unix philosophy

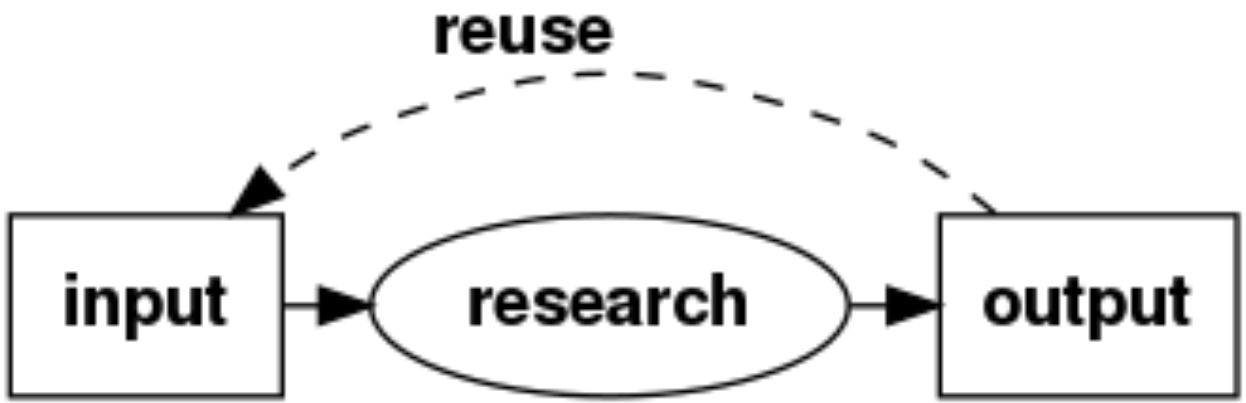
Expect the output of every program to become the input to another, as yet unknown, program.

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.


reuse
input
computer program
output

a research philosophy

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


reuse
input
research
output

February 24, 2020

DOI 10.5281/zenodo.8475 Dataset Open Access

Terrestrial Parasite Tracker indexed biotic interactions and review summary

Poelen, Jorrit H.; Seltmann, Katja C.

Terrestrial Parasite Tracker indexed biotic interactions and review summary.

The Terrestrial Parasite Tracker (TPT) project began in 2019 and is funded by the National Science foundation to mobilize data from vector and ectoparasite collections to data aggregators (e.g., iDigBio, GBIF) to help build a comprehensive picture of arthropod host-association evolution, distributions, and the ecological interactions of disease vectors which will assist scientists, educators, land managers, and policy makers. Arthropod parasites often are important to human and wildlife health and safety as vectors of pathogens, and it is critical to digitize these specimens so that they, and their biotic interaction data, will be available to help understand and predict the spread of human and wildlife disease.

This is the first archive of TPT data indexed by Global Biotic Interactions (GloBI). GloBI provides open access to finding species interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing open datasets using open source software.

Funding:
The creation of this archive was made possible by the National Science Foundation award "Collaborative Research: Digitization TCN: Digitizing collections to trace parasite-host associations and predict the spread of vector-borne disease," Award numbers DBI:1901932 and DBI:1901926

https://globalbioticinteractions.org/parasitetracker

Establish a citable baseline.

Events

1 Oct 2019 - Parasite Tracker kick-off presentation / video.

24 Feb 2020 - Publication of `Poelen, Jorrit H., & Seltmann, Katja. (2020). Terrestrial Parasite Tracker indexed biotic interactions and review summary (Version 0.1) [Data set]. Zenodo. [http://doi.org/10.5281/zenodo.3685365](http://doi.org/10.5281/zenodo.3685365)` . For summaries see indexed_interactions_by_collection.tsv and review_summary_by_collection.tsv. For full report with review comments, indexed interactions and source datasets, please visit the Zenodo data publication at http://doi.org/10.5281/zenodo.3685365. **If you'd like to be added as a co-author, please add yourself at https://pad.carpentries.org/tpt-datapaper-authors by the end of February or contact Jorrit or Katja directly.**

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>

2020 - Jorrit attends TPT workshop at Field Museum, Chicago

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.

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

collected_at

classified_as

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.

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.

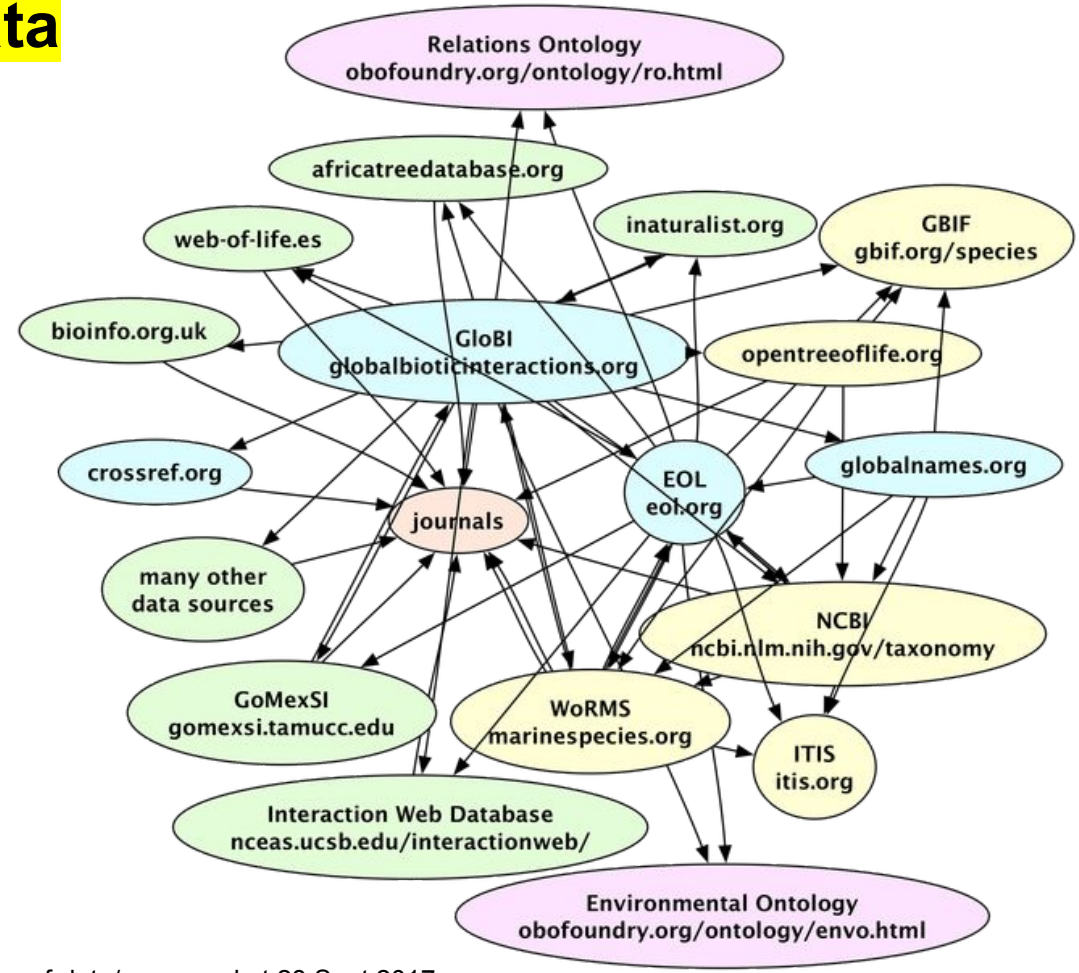


Figure from http://globalbioticinteractions.org/2016/02/08/ecology-of-data/ accessed at 28 Sept 2017

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.

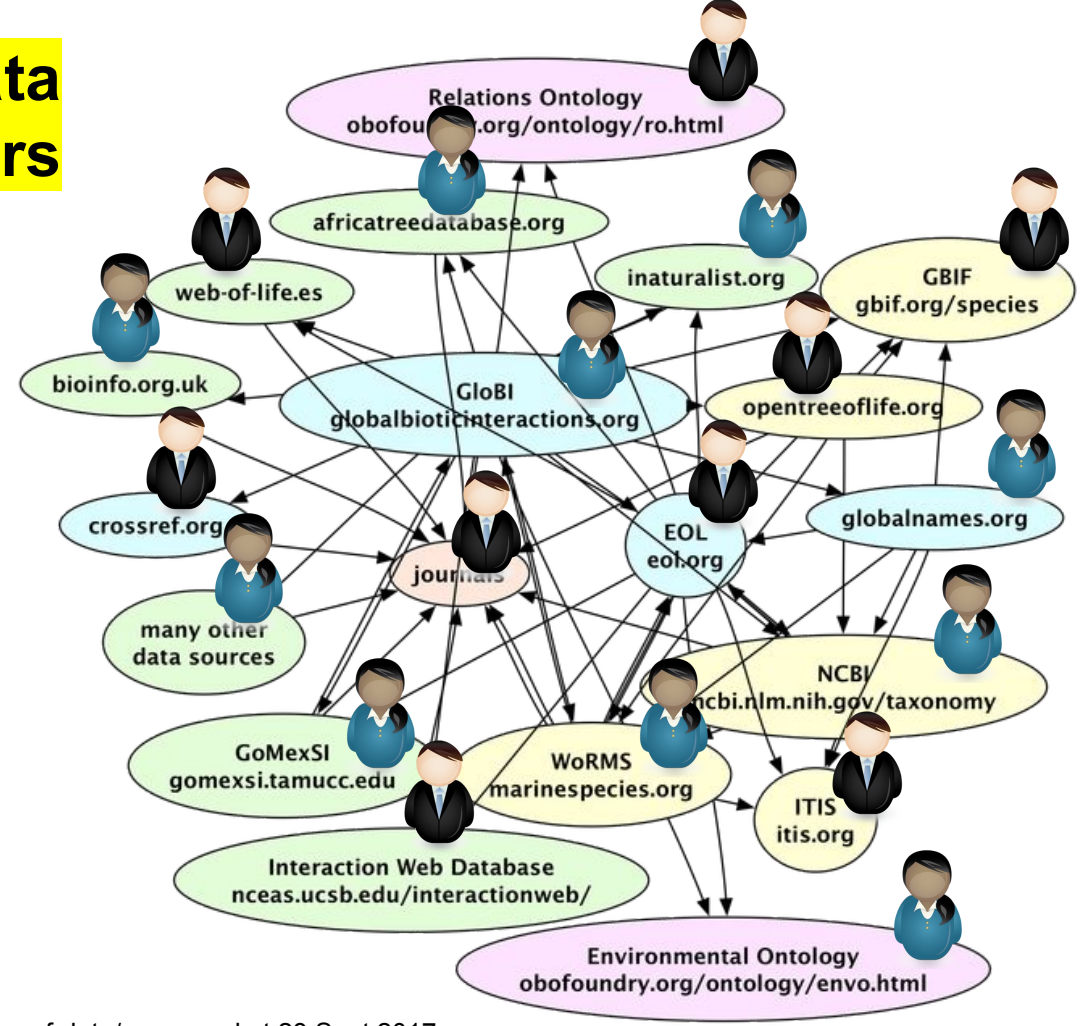
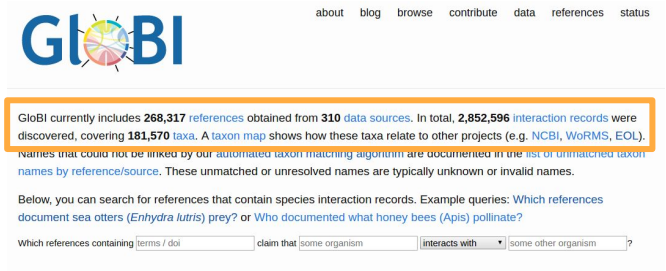
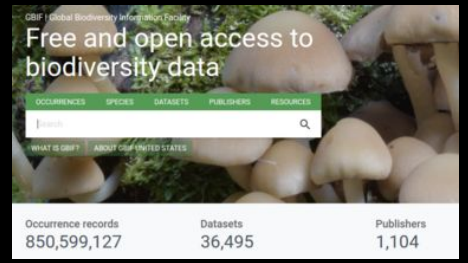


Figure from http://globalbioticinteractions.org/2016/02/08/ecology-of-data/ accessed at 28 Sept 2017



http://globalbioticinteractions.org/references
Accessed at 28 Sept 2017

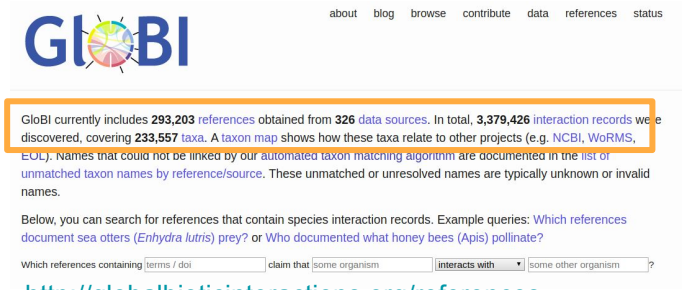


https://gbif.org
Accessed at 28 Sept 2017

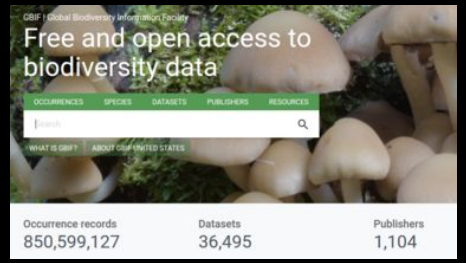
2.8M records	950.6M records
0.1k datasets	36.5k datasets
~100k **taxa**	~1-2M **species**

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.



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

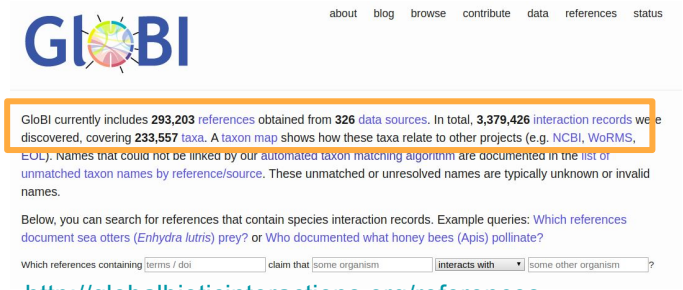


https://gbif.org
Accessed at 26 Feb 2018

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

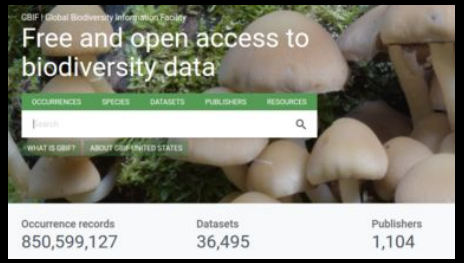
Eltonian shortfall*: a lack of species-interaction records

*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.



http://globalbioticinteractions.org/references
Accessed at 9 Jan 2020

GBIF Global Biodiversity Information Facility



https://gbif.org
Accessed at 9 Jan 2020

4.4M records	**1.4B records**
0.2k datasets	50.0k datasets
~0.2M **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.

MENU SCIENTIFIC REPORTS

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

ECOGRAPHY

AN OPEN ACCESS JOURNAL FROM 1 JANUARY 2020

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

ANNUAL REVIEWS

JOURNALS A-Z | JOURNAL INFO | PRICING & SUBSCRIPTIONS | SUBS

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

Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity

Annual Review of Ecology, Evolution, and Systematics
Vol. 46:523-549 (Volume publication date December 2015)
First published online as a Review in Advance on October 28, 2015
https://doi.org/10.1146/annurev-ecolsys-112414-054400

Joaquín Hortal,[1,2,3,*] Francesco de Bello,[4,5] José Alexandre F. Diniz-Filho,[2] Thomas M. Lewinsohn,[6] Jorge M. Lobo,[1] and Richard J. Ladle[7,8,*]

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 Sassanfar, Yan-Jiun Lee, Peter Weigele, 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

topics

> Vision and Context

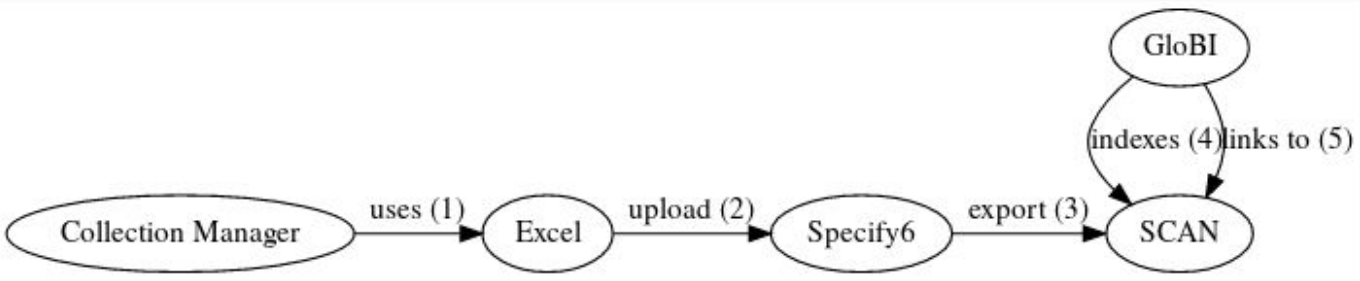
> Integration Profiles

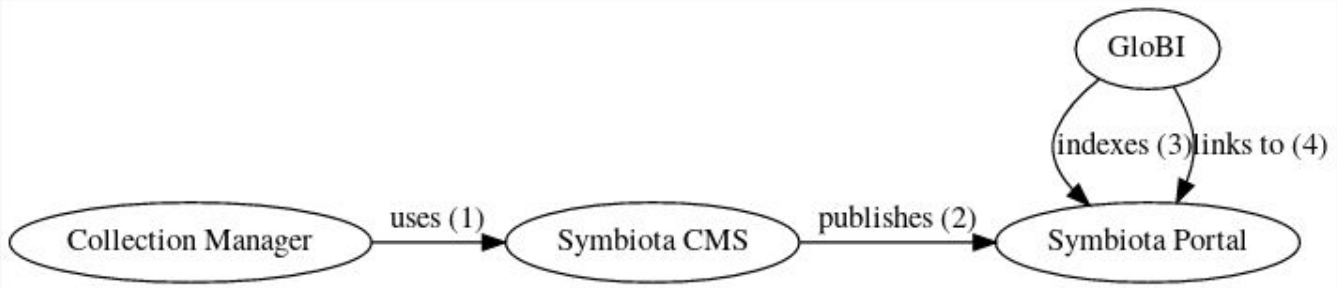
> Reviews and Indexes

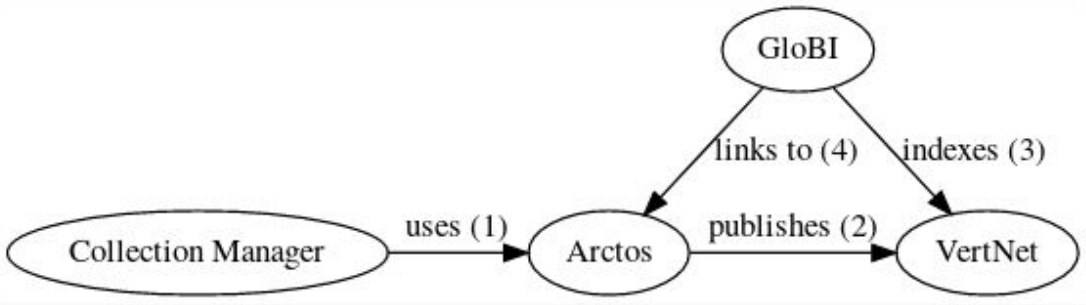
> Interaction Terms and Translation Tables

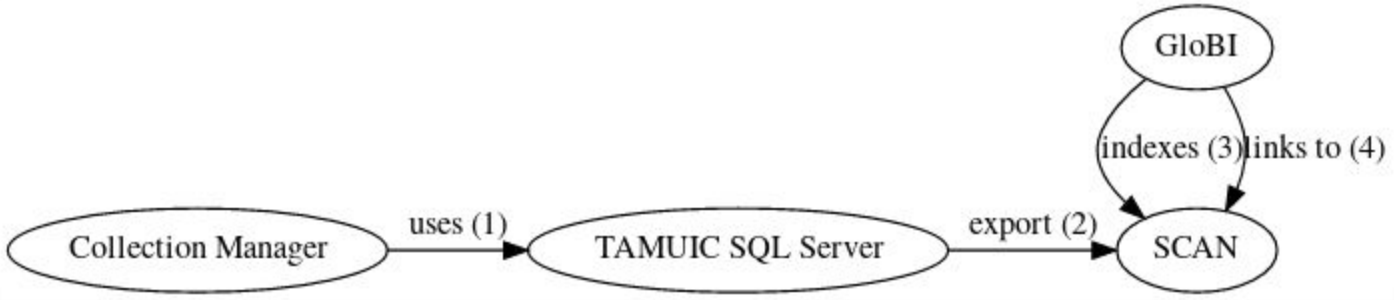
https://globalbioticinteractions.org/parasitetracker

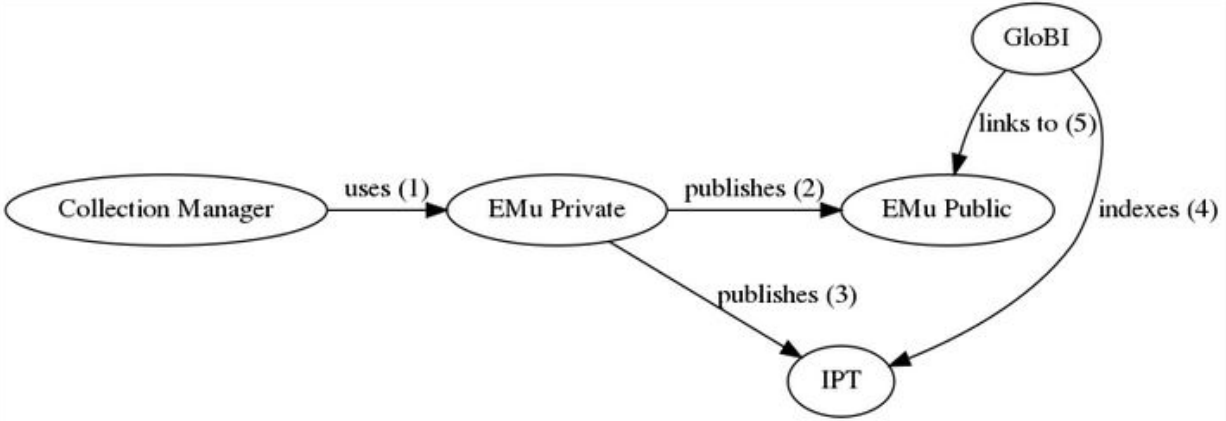
Integration Profiles: Data Workflows

Specify	GloBI Integration Profile
authors	Ralph Holzenthal (UMSP), Robin Thomson (UMSP)
actors	Collection Manager, Excel, Specify6, https://scan-bugs.org
integration method	Collection Manager enters records in excel, then uploads the records in batch into Specify. A manual export to https://scan-bugs.org is done periodically to provide updates to GBIF, iDigBio and GloBI. If updates are made to existing Specify records, a new batch export is needed to update https://scan-bugs.org .
diagram	
example collection	
open questions	1. Which darwin core archive field to use for assocations? associatedTaxa, associatedOccurrences, Resource Relationship extension. 2. Can we automate the export of Specify records to SCAN?
references	Linking Specify data to SCAN Collection Provided by Laura Prado, U. of Wisconsin (December 8, 2018) https://scan-all-bugs.org/?page_id=2084

Symbiota	GloBI Integration Profile
authors	Katja Seltmann (UCSB)
actors	Collection Manager, Symbiota CMS[1], Symbiota Portal[1], GloBI
integration method	A collection manager uses the “associatedTaxa” fields in Symbiota CMS to record host-parasite associations. The Symbiota CMS periodically publishes their data to a Symbiota Portal (e.g., https://scan-bugs.org ,) . After successful publication, the Symbiota Portal includes the updates data archive in their list of available datasets through their RSS feed. GloBI indexes all data archives in the list of available datasets. For each dataset, GloBI looks for association records in associatedTaxa, associatedOccurrences, dynamicProperties field as well as Resource Relationship and Associated Taxa Extensions. In this case, only associatedTaxa fields are encountered and related records are indexed accordingly.
diagram	
example collection	UCSB-IZC

Arctos	GloBI Integration Profile
authors	Dusty (Arctos dev), Mariel Campbell (MSB), Teresa Mayfield-Meyer (MSB)
actors	Arctos, VertNet, GloBI
integration method	A collection manager uses Arctos to establish associations or relationships between records. Arctos periodically shares data with VertNet. VertNet uses GBIF IPT software to publish data archives. VertNet publishes a list of available datasets in the form of a RSS feed, including those shared by Arctos. Periodically, GloBI finds, and downloads, Arctos related data archives in VertNet. Then, GloBI indexes the associatedOccurrences fields of records in these Arctos data archives. The associatedOccurrences contain the association type (e.g., “eats”) and a pointer to the occurrence id of the linked record.
diagram	
example collection	MSB-PARA

TAMUIC	GloBI Integration Profile
authors	John Oswald
actors	Collection Manager, SQL Server, SCAN, GloBI
integration method	A collection manager uses SQL Server to manage collection record relations. The collection manager episodically exports datasets from SQL Server to SCAN as a Darwin Core Archive. Periodically, GloBI indexes TAMUIC related datasets as part of indexing all of SCAN.
diagram	

EMu/GBIF	GloBI Integration Profile
authors	Kate Webbink, Janeen Jones
actors	FMNH Collection Manager, EMu, IPT, GloBI
integration method	A collection manager uses EMu to establish associations or relationships between Catalogue (occurrence) records. The collection manager periodically exports datasets from EMu, and the IT Department publishes those datasets as resources on the FMNH IPT - fmipt.fieldmuseum.org. Datasets ("IPT resources") that include interactions among occurrences will include a Darwin Core "Resource Relationship" extension. The fmipt site publishes a list of available IPT resources in the form of a RSS feed (https://fmipt.fieldmuseum.org/ipt/rss.do), similarly to VertNet. Periodically, GloBI could find and download FMNH-related data archives in the fmipt. Then, GloBI could index the resourceID fields of records in these FMNH data archives (IPT resources). The relationshipOfResource field contains the relationship type (e.g., "stomach contents of") and the relatedResourceID field contains a pointer to the occurrenceID of the linked record.
diagram	A workflow using a public EMu website for GloBI to link back into: 

Specify	GloBI Integration Profile
authors	Erika Tucker (UMMZ), Barry Oconner (UMMZ)
actors	Collection Manager, Specify, https://gbif.org/
integration method	Collection Manager enters records in Specify. A manual csv export of full Specify database (including host records) is shared with, and indexed by, GloBI. If updates are made to existing Specify records, a manual export performed and GloBI is notified.
diagram	
example collection	
open questions	1. Which darwin core archive field to use for assocations? associatedTaxa, associatedOccurrences, Resource Relationship extension. 2. How to establish an automated the export of Specify records via (UMich) IPT? 3. How to best link to individual specimen records? 4. How to best cite specimen records? 5. How to establish reliable links to non-UMMZI host records?
references	2020-01-23 Meeting notes and Jan 2020 email exchanges between Jorrit (GloBI), Erika (UMMZ), Barry (UMMZ)

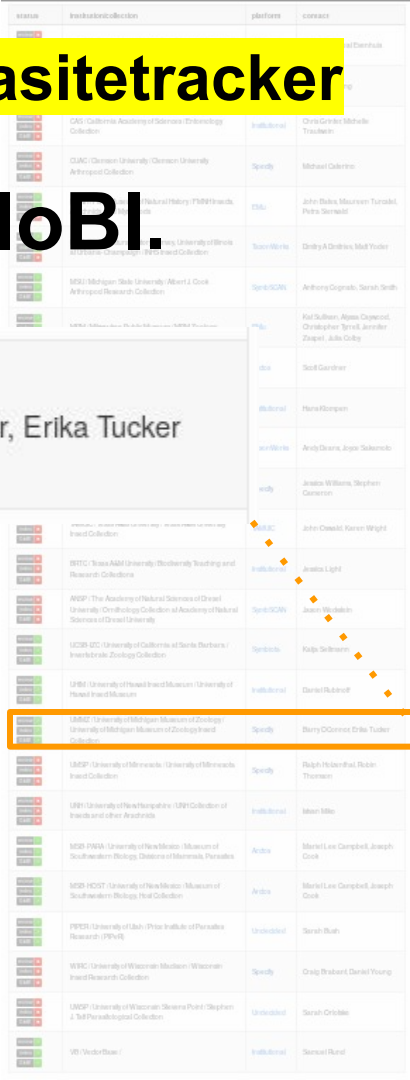
topics

> Vision and Context
> Integration Profiles
> Reviews and Indexes
> Interaction Terms and Translation Tables

https://globalbioticinteractions.org/parasitetracker

Collection records as seen by GloBI.

review ✓ index ✓ GloBI ✓	UMMZ / University of Michigan Museum of Zoology / University of Michigan Museum of Zoology Insect Collection	Specify	Barry OConnor, Erika Tucker



https://globalbioticinteractions.org/parasitetracker

Collection records as seen by GloBI.

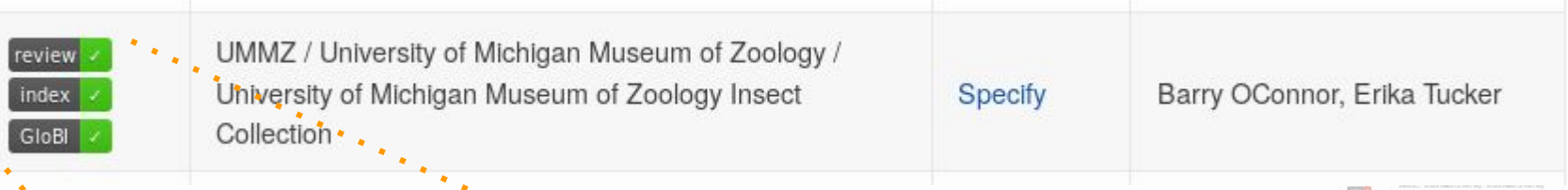
review ✓ index ✓ GloBI ✓	UMMZ / University of Michigan Museum of Zoology / University of Michigan Museum of Zoology Insect Collection	Specify	Barry OConnor, Erika Tucker

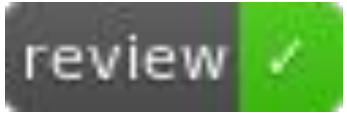
review ✓

click to get review notes

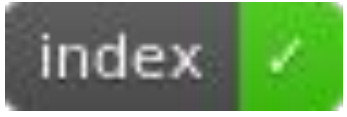
https://globalbioticinteractions.org/parasitetracker

Collection records as seen by GloBI.





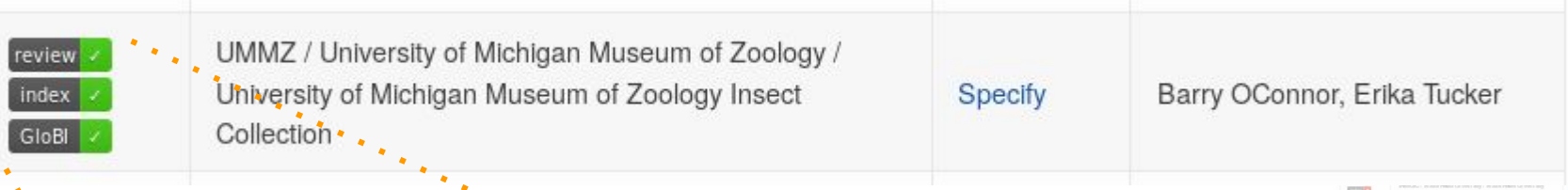
click to get review notes

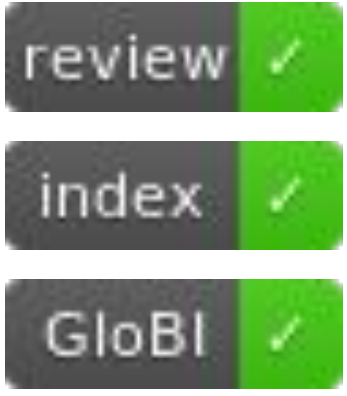


or ... download interaction index

https://globalbioticinteractions.org/parasitetracker

Collection records as seen by GloBI.





click to get review notes

or ... download interaction index

and to browse at least 1 record

topics

> Vision and Context
> Integration Profiles
> Reviews and Indexes
> Interaction Terms and Translation Tables

https://globalbioticinteractions.org/parasitetracker

Interaction Terms / Translation Tables

Supported Terms

Translation Tables

GloBI uses a subset of biotic interaction (or association) terms defined in the OBO Relations Ontology (OBO RO) to help classify and index biotic associations in collection records. Verbatim association types (e.g., `on`, `parasite of`, `found on`) are explicitly mapped into these OBO RO terms using translation tables. GloBI keeps a default translation tables and specific collection may choose to provide their own (see e.g., INHS-Insects).

resource	description
OBO Relations Ontology project page	OBO RO contains many kinds of terms, not just biotic associations terms
List of OBO RO Biotic Interaction Terms with definitions	a table of RO biotic interaction terms and their definitions (if available)
List of GloBI Supported Interaction Terms	subset of RO interactions terms that GloBI uses for indexing
Default Verbatim Terms Translation Table	the translation table used by GloBI to maps verbatim interaction terms to supported interaction terms
Example of Custom Verbatim Terms Translation Table	if provided/needed, GloBI can use a custom mapping provided by a collection

The OBO RO is far from complete and we expect to add new terms and improve definitions as needed. Also, GloBI translation tables can be easily updated when needed. Please open an issue if you have questions or suggestions.

https://globalbioticinteractions.org/parasitetracker

Interaction Terms / Translation Tables

Supported Terms

Translation Tables

GloBI uses a subset of biotic interaction (or as
index biotic associations in collection records.
mapped into these OBO RO terms using trans
choose to provide their own (see e.g., INHS-Ir

resource
OBO Relations Ontology project page
List of OBO RO Biotic Interaction Terms with definitions
List of GloBI Supported Interaction Terms
Default Verbatim Terms Translation Table
Example of Custom Verbatim Terms Translation Table

	interaction_type_label	interaction_type_id
2	preysOn	http://purl.obolibrary.org/obo/RO_0002439
3	parasiteOf	http://purl.obolibrary.org/obo/RO_0002444
4	hasHost	http://purl.obolibrary.org/obo/RO_0002454
5	interactsWith	http://purl.obolibrary.org/obo/RO_0002437
6	hostOf	http://purl.obolibrary.org/obo/RO_0002453

subset of RO interaction terms that GloBI uses for indexing

the translation table used by GloBI to maps verbatim interaction terms to supported interaction terms

if provided/needed, GloBI can use a custom mapping provided by a collection

The OBO RO is far from complete and we expect to add new terms and improve definitions as needed. Also, GloBI translation tables can be easily updated when needed. Please open an issue if you have questions or suggestions.

https://globalbioticinteractions.org/parasitetracker

Interaction Terms / Translation Tables

provided_interaction_type_label	provided_interaction_type_id	mapped_to_interaction_type_label	mapped_to_interaction_type_id
associated with		interactsWith	http://purl.obolibrary.org/obo/RO_0002437
ex		hasHost	http://purl.obolibrary.org/obo/RO_0002454
ex.		hasHost	http://purl.obolibrary.org/obo/RO_0002454
reared ex		hasHost	http://purl.obolibrary.org/obo/RO_0002454
reared ex		hasHost	http://purl.obolibrary.org/obo/RO_0002454

with definitions — a table of RO biotic interaction terms and their definitions (if available)

List of GloBI Supported Interaction Terms — subset of RO interactions terms that GloBI uses for indexing

Default Verbatim Terms Translation Table — the translation table used by GloBI to maps verbatim interaction terms to supported interaction terms

Example of Custom Verbatim Terms Translation Table — if provided/needed, GloBI can use a custom mapping provided by a collection

The OBO RO is far from complete and we expect to add new terms and improve definitions as needed. Also, GloBI translation tables can be easily updated when needed. Please open an issue if you have questions or suggestions.

topics

> Vision and Context
> Integration Profiles
> Reviews and Indexes
> Interaction Terms and Translation Tables

https://globalbioticinteractions.org
info@globalbioticinteractions.org
@GlobalBiotic

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.

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), the Smithsonian Institution (SI) (T15CC10297-002, 2016) and NSF:DBI:1901932 .

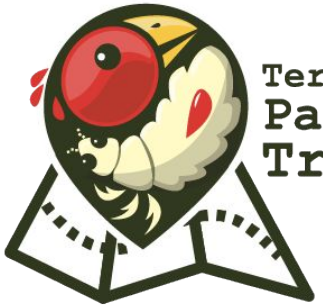
partners

GloBI

TEXAS A&M UNIVERSITY-CORPUS CHRISTI
GoMexSI
CENTER FOR COASTAL STUDIES

eol
Encyclopedia of Life

Heitlinger lab
Ecology
Evolution
molecular Mechanism
Interaction
parasites
of host and


Terrestrial Parasite Tracker

How can **we** help **each other** to share **our** parasite-host records?