

Pensoft - Naturalis - GloBI meetings

❖ 12/10/2021

Attendees:

- Mariya Dimitrova
- Lyubomir Penev
- Teodor Georgiev
- Georgi Zhelezov
- Niki Kyriakopoulou
- Jorrit Poelen

Meeting notes:

Notes from previous meetings between GloBI and Pensoft:

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

How to access GloBI data in graph format?

- SPARQL endpoint (now running):
https://lod.globalbioticinteractions.org/globi/sparql?query=SELECT+*+WHERE+%7B%3FX+%3FP+%3FY%7D+LIMIT+10&output=json&stylesheet=%2Fxml-to-html.xsl&force-accept=text%2Fplain
- Neo4j: <https://github.com/globalbioticinteractions/globalbioticinteractions/wiki/Cypher>
- Rglobi library in R to explore globi data: <https://github.com/ropensci/rglobi>
- GloBI API: <https://github.com/globalbioticinteractions/globalbioticinteractions/wiki/API>

Graphs of interactions can also be downloaded and run locally (they are not so many triples):

<https://globalbioticinteractions.org/data>

Datasets indexed by GloBI can be found here:

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

Niki: Are there any datasets with specimen data?

Jorrit: Yes, see Terrestrial Parasites Tracker

(<https://www.globalbioticinteractions.org/2020/03/05/tracking-parasites/>)

Also, the EOL has implemented a rule-based approach to “filter out” false positive interactions (e.g. plants pollinating insects and such). Their work contributes to make GloBI more accurate by providing ‘refuted by’ statements.

More info:

https://github.com/eliagbayani/eol_php_code/blob/master/lib/connectors/GloBIDataAPI.php

https://github.com/eliagbayani/eol_php_code/blob/master/lib/connectors/Globi_Refuted_Records.php

With some narrative via poster:

Schulz, Katja, Hammock, Jennifer, Poelen, Jorrit H, & Agbayani, Eli. (2020). Management of Biotic Interaction Data in the Encyclopedia of Life (0.1). Digital Data in Biodiversity

Research, Indiana University, Bloomington, Indiana. Zenodo.
<https://doi.org/10.5281/zenodo.4015329>

There are also interactions with specimen data from herbaria in GloBI which have some fuzzy relationships.

Mariya: What taxa/interaction types would Naturalis like to focus on in BiCIKL?

Niki: Pollination interactions.

Jorrit: There are some really interesting projects about pollinator interactions:

- Big Bee: <http://big-bee.net/> (Started recently); Seltnann, K. et al., 2021. Announcing Big-Bee: An initiative to promote understanding of bees through image and trait digitization. Biodiversity Information Science and Standards, 5. Available at: <https://doi.org/10.3897/biss.5.74037>.
- San Diego Zoo pollinator project: <https://saveplants.org/pollinator-search/>

Mariya: What sort of ML algorithm is Naturalis looking to build? Predicting probability that 2 species interact or predicting the second participating species in an interaction (Bees pollinate ... ?)

How can we use the existing OpenBiodiv database as a source of knowledge?

Suggestion:

- Build ML algorithm based on existing GloBI data on pollination (data is not perfect, may include false positives etc)
- Use OpenBiodiv as a data source for validation. E.g. each treatment section contains information about the specimens mentioned in it. We can use this information to try and predict (using the algorithm) whether each 2 species participate in a pollination interaction based on their co-occurrence in the same treatment section. This narrows down the choice of participating taxa in the interaction a lot so it may not be a “sound” ML approach but it makes sense as a use case. Also, we have the texts for the treatments so a human can easily check whether the prediction is correct.

Lyubo: Can also use the Pensoft Annotator if the ML algorithm requires additional information about presence of certain terms (morphology terms from morphology ontologies, habitats, etc).