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]

[1]Encyclopedia of Life, Smithsonian National Museum of Natural History, Washington, DC, USA
[2]Global 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.
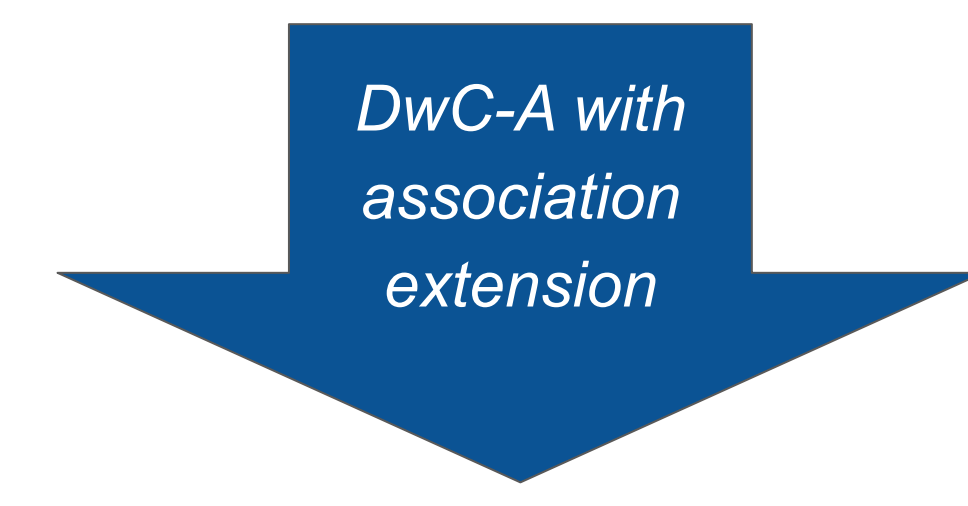
Trophic Web
- Predator
- Competitor
- Prey
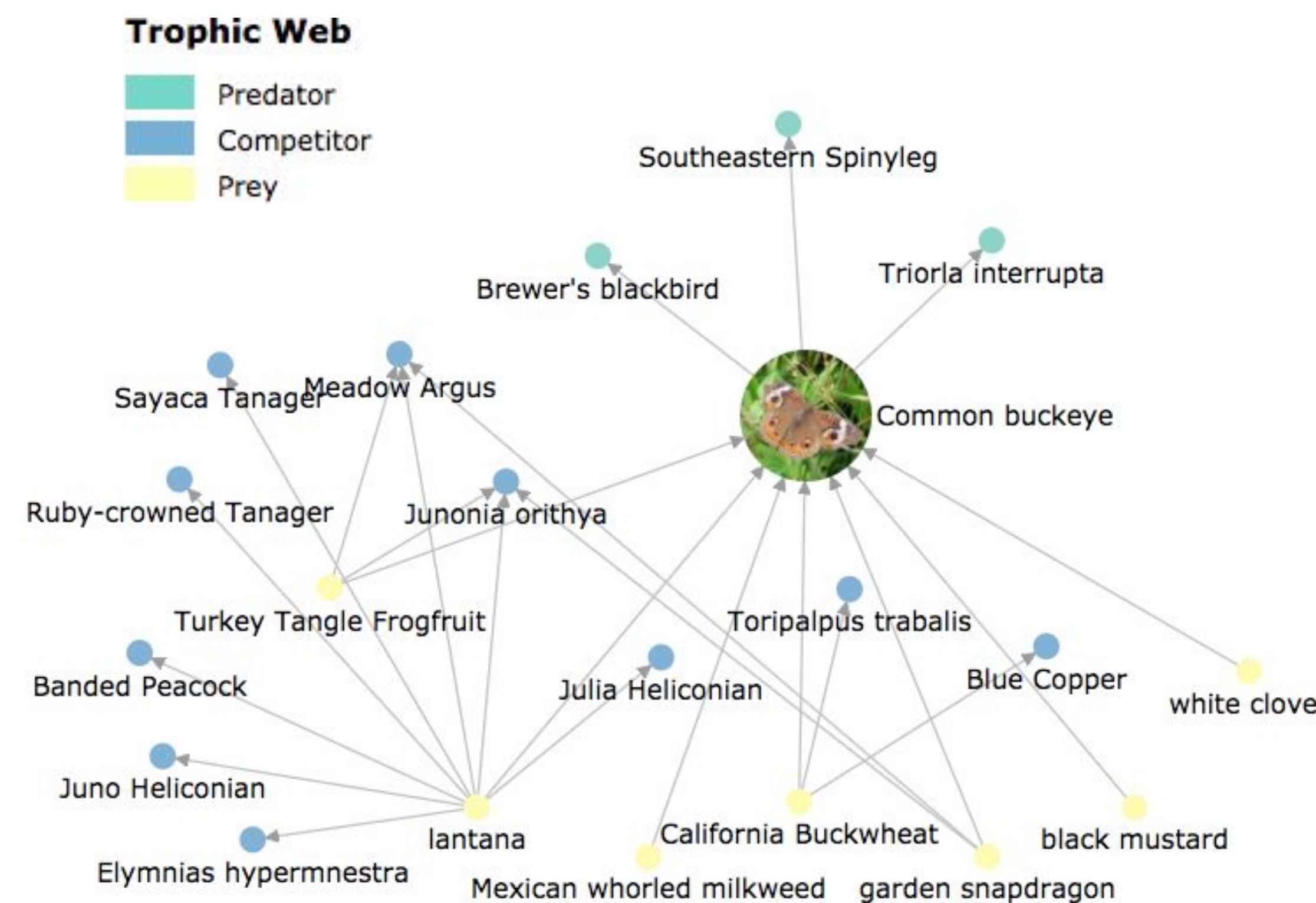
639 data sources • **4,891,426** interaction records • **331,512** taxa

scientific literature | databases | natural history collections | citizen science projects | text mining

DwC-A with association extension

interaction records refuted by EOL

EOL data validation	
recode (1.3 % of records)	**reject** (0.4 % of records)
imprecise association claims, e.g., most "pollinated by" interaction records are based on flower visitation, not evidence of pollination.	invalid association claims, e.g., records asserting animals being "eaten by" plants that are not on a carnivorous plant whitelist.

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*