

# taxize: taxonomic search and retrieval in R

Eduard Szöcs<sup>a</sup> and Scott A. Chamberlain<sup>b</sup>

<sup>a</sup>Institute for Environmental Sciences  
University Koblenz-Landau  
Landau, Germany  
eduardsoecs@googlemail.com

<sup>b</sup>Biology Department  
Simon Fraser University  
Burnaby, Canada  
myrmecocystus@gmail.com

**Keywords:** big data; ...;

**Abstract:** Ecologists use data from many sources (e.g. own sampling, governmental monitoring data, data from web-databases, ...). All these data have a common ground: the taxonomic names of species they contain. For a data analysis a crucial data-cleaning step is to match species identities across different datasets Boyle et al. (2013). Taxonomic data is becoming increasingly available on the web, but scientists need an easy and reproducible way to access it.

We present *taxize* Chamberlain and Szöcs (2013) - an open-source software package for the R language (freely available from <http://cran.r-project.org/web/packages/taxize/index.html>). Taxize is part of the rOpenSci Initiative (<http://ropensci.org/>) and currently under development. It provides simple, programmatic access to taxonomic data for 14 data sources around the web. It simplifies data-cleaning steps, like fixing taxonomic names, aggregating data to a specific taxonomic level and matching tables by species names with different taxonomic resolutions.

We present this taxonomic tool belt in R and outline a suite of use cases for which *taxize* is ideally suited. The *taxize* package facilitates open and reproducible science by allowing taxonomic data collection to be done in the open-source R platform. We hope *taxize* will be taken up by the community and developed collaboratively.

## References

- Boyle, B., Hopkins, N., Lu, Z., Garay, J. A. R., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., Mckay, S. J., Lowry, S., Freeland, C., Peet, R. K., and Enquist, B. J. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, 14(1):16.
- Chamberlain, S. A. and Szöcs, E. (2013). *taxize: taxonomic search and retrieval in r* [v2; ref status: indexed, <http://f1000r.es/24v>]. *F1000Research*, 2(191).