




neotoma2: An R package to access data from the Neotoma Paleocology Database

Socorro Dominguez Vidaña ¹ and Simon J Goring ^{2,3}

1 HT Data 2 Department of Geography, University of Wisconsin – Madison 3 Center for Climatic Research, University of Wisconsin – Madison

DOI: [10.21105/joss.05561](https://doi.org/10.21105/joss.05561)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Chris Vernon](#) 

Reviewers:

- [@kanishkan91](#)
- [@mhesselbarth](#)

Submitted: 03 May 2023

Published: 28 November 2023

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The `neotoma2` R package is a tool to access and manipulate data from the Neotoma Paleocology Database (<https://www.neotomadb.org>) within the R environment. Neotoma is a community curated paleoecological data resource ([Williams et al., 2018](#)), containing nearly 9 million unique observations of paleoecological proxies with global coverage from 37 constituent databases. The package uses the Neotoma API v2.0 ([Simon J. Goring, 2023](#)) as a tool to import records from the Neotoma database, allowing researchers to examine taxonomic, spatial and temporal patterns across space and time over the last 5.4 million years. The R package allows researchers to download, and create new records using `get` and `set` functions (e.g., `get_sites()`, `set_sites()`) respectively. This provides researchers with the opportunity to develop dynamic workflows that include data generated locally, and not yet submitted to the Neotoma database. The `set` functions are intended as a precursor to utilities to upload data directly to Neotoma, although this functionality is not yet available.

The `neotoma2` R package has been under dynamic development for over a year, but has been used for teaching and training ([Simon J. Goring & Dominguez Vidaña, 2023](#)). This release of the `neotoma2` R package is a clean release of the package, with all of the core features provided and extensive test coverage implemented.

Statement of Need

The `neotoma` R package ([S. Goring et al., 2015](#)) leveraged the Neotoma Paleocology Database v1.0 API and had been one of the primary tools for researchers working with data from Neotoma ([Byun et al., 2021](#); [Kujawa et al., 2016](#); [Wang et al., 2023](#)). Changes to the underlying database and a rebuilding of the API required new data objects within the R package to more closely align to the Neotoma data model ([Grimm, 2008](#)). Additionally, the original v1.0 API that was accessed by the `neotoma` package was deprecated in 2020, meaning the `neotoma` package could no longer access data from Neotoma.

The broad user community for Neotoma ([Simon James Goring et al., 2018](#); [Williams et al., 2018](#)) requires a toolset that can access and manage data for each of the more than 40 dataset types within Neotoma and so extensive metadata must be accessed for each record. The `neotoma2` package conforms to a `tidyverse` ([Wickham et al., 2019](#)) approach for data management, with methods and data objects that are suited to piping using the `%>%` (or `|>`) pipe convention and with implementations for filtering and other common `dplyr` methods. The package also now uses “long” `data.frames` and `tibbles` by default, using the `toWide()` function to transform data into “wide” tables for use with common ecological data packages such as `vegan` ([Oksanen et al., 2022](#)). Data objects in the `neotoma2` package now more closely resemble the underlying data model within Neotoma (<https://open.neotomadb.org/dbschema>) than in the previous package. Most importantly the `neotoma2` package provides a toolset for

paleoecologists, ecologists, conservation ecologists, archaeologists, and others, to access and examine the broad range of fossil data contained within the Neotoma Paleocology Database.

Data from the Neotoma Database can be accessed through the public API (<https://api.neotomadb.org>) or a PostgreSQL database snapshot with a database client [Williams et al. (2018); <https://www.neotomadb.org/data/db-snapshots>] or through the EarthLife Consortium API (Uhen et al., 2021). The R package will simplify many of the operations required to assemble and manipulate datasets, and provides functions that support researchers as part of their analytic workflows by linking Neotoma directly to packages within R used for ecological and earth science research, data visualization and statistical analysis.

Acknowledgements

We acknowledge contributions from the Neotoma Paleocology Community, the participants of our workshops for the European Pollen Database, American Quaternary Association and the International Association of Limnologists/International Paleolimnology Association, and members of the EarthCube community. This work was funded through a grant to SJG from the National Science Foundation (NSF-1948926).

References

- Byun, E., Sato, H., Cowling, S. A., & Finkelstein, S. A. (2021). Extensive wetland development in mid-latitude North America during the Bølling–Allerød. *Nature Geoscience*, *14*(1), 30–35. <https://doi.org/10.1038/s41561-020-00670-4>
- Goring, Simon J. (2023). Neotoma Paleocology Database API. In *GitHub repository*. GitHub. <https://doi.org/10.5281/zenodo.7893836>
- Goring, Simon J., & Dominguez Vidaña, S. (2023). Neotoma workshops APD release. In *GitHub repository*. GitHub. <https://doi.org/10.5281/zenodo.7849972>
- Goring, Simon James, Graham, R., Loeffler, S., Myrbo, A., Oliver, J. S., Ormond, C., & Williams, J. W. (2018). *The Neotoma Paleocology Database: A research outreach nexus*. Cambridge University Press. <https://doi.org/10.1017/9781108681582>
- Goring, S., Dawson, A., Simpson, G., Ram, K., Graham, R., Grimm, E., & Williams, J. W. (2015). neotoma: A programmatic interface to the Neotoma Paleocological Database. *Open Quaternary*, *1*(1). <https://doi.org/10.5334/oq.ab>
- Grimm, E. (2008). Neotoma: An ecosystem database for the Pliocene, Pleistocene, and Holocene. *Illinois State Museum Scientific Papers E Series*, *1*.
- Kujawa, E. R., Goring, S., Dawson, A., Calcote, R., Grimm, E. C., Hotchkiss, S. C., Jackson, S. T., Lynch, E. A., McLachlan, J., St-Jacques, J.-M., Umbanhowar Jr., C., & Williams, J. W. (2016). The effects of anthropogenic land cover change on pollen-vegetation relationships in the American Midwest. *Anthropocene*, *15*, 60–71. <https://doi.org/10.1016/j.ancene.2016.09.005>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., ... Weedon, J. (2022). *Vegan: Community ecology package*. <https://CRAN.R-project.org/package=vegan>
- Uhen, M. D., Buckland, P. I., Goring, S. J., Jenkins, J. P., & Williams, J. W. (2021). The EarthLife Consortium API: An extensible, open-source service for accessing fossil data and taxonomies from multiple community paleodata resources. *Frontiers of Biogeography*, *13*(2). <https://doi.org/10.21425/F5FBG50711>

- Wang, Y., Pineda-Munoz, S., & McGuire, J. L. (2023). Plants maintain climate fidelity in the face of dynamic climate change. *Proceedings of the National Academy of Sciences*, *120*(7), e2201946119. <https://doi.org/10.1073/pnas.2201946119>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>
- Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabres, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., ... Takahara, H. (2018). The Neotoma Paleocology Database, a multiproxy, international, community-curated data resource. *Quaternary Research*, *89*(1), 156–177. <https://doi.org/10.1017/qua.2017.105>