



Neotoma Paleoecology Database: Recent Updates

By Jack Williams, on behalf of the Neotoma Paleoecology Database Consortium
University of Wisconsin-Madison, jwwilliams1@wisc.edu

The mission of the Neotoma Paleoecology Database is to serve the Quaternary community and support global-change research by providing an open, high-quality, sustainable, community-curated repository for multiple kinds of paleoecological data. Since the last Quaternary Times report (May, 2017), Neotoma has continued to grow, in multiple ways: data volume, supported data types and constituent databases, data services, supported science and publications, and, most importantly, in its community of data contributors, data stewards, and leaders.

Data Volume

As of November 3, 2019, Neotoma now holds over 6.9 million individual observations, from 37,732 datasets, distributed among 18,212 sites. The global scope of Neotoma is increasing, particularly with the current data mobilization campaigns for the European Pollen Database and Latin American Pollen Database, both constituent databases in Neotoma. A new campaign has just been launched for the African Pollen Database, with funding from the Belmont Forum (lead PI: Nick McKay) and a workshop in Oct 2019 hosted by the Research Institute for Development (IRD).

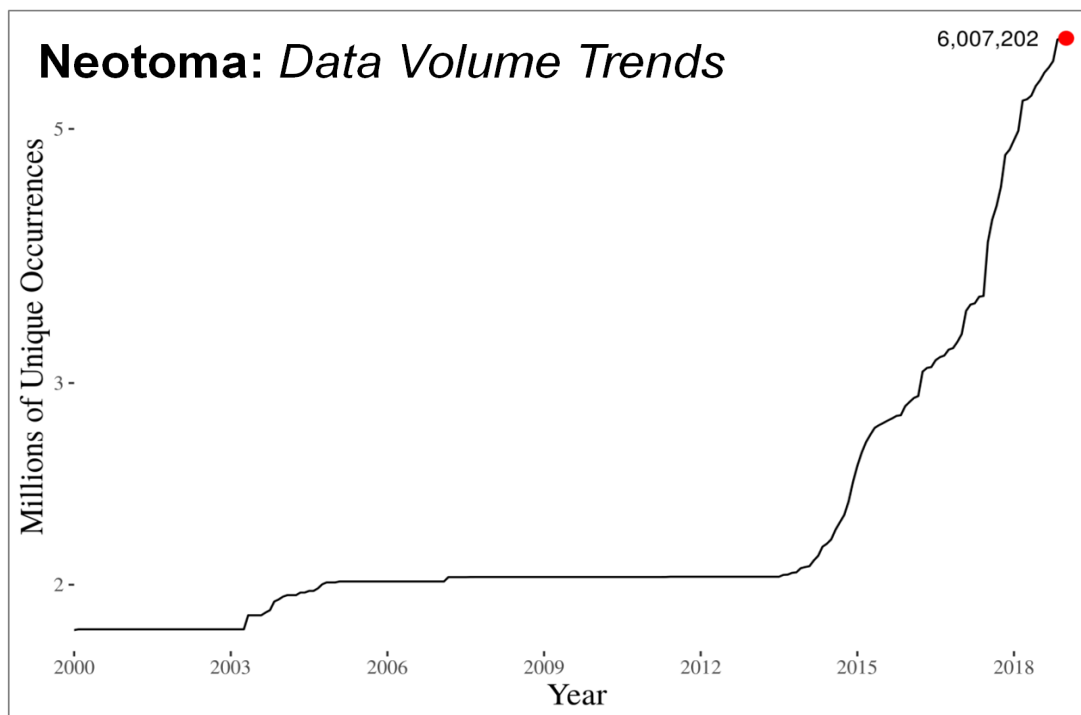


Figure: S.Goring

Supported Data Types and Constituent Databases

Neotoma now has sizeable holdings in diatoms (1,149 stratigraphic datasets, 1,305 surface samples), insects (210 stratigraphic datasets, 13 surface samples), ostracodes (12 stratigraphic records, 5,057 surface samples), pollen (4,567 stratigraphic datasets, 3,166 surface samples), testate amoebae (85 stratigraphic datasets, 2,393 surface samples), vertebrates (5,028 stratigraphic datasets), and water chemistry (7,655 datasets). Neotoma also stores over 5,700 geochronological datasets, most linked to age-depth models. Other supported data types include biomarkers, chironomids, cladocerans, loss-on-ignition, macroinvertebrates, phytoliths, plant macrofossils, stable isotopes, XRF, and XRD. Each data type is associated with one or more Constituent Databases, each with one or more expert Data Stewards.

Supported Science and Publications

Neotoma is designed to support broad-scale, multi-site, and multi-proxy research. Current research collaborations include ACCEDE, CLIMATE12K,

ECORE3, HOPE, LANDCOVER6K, and SKOPE. In these partnerships, researchers building large data syntheses benefit from the large curated data resources. In return, macro-scale researchers can improve Neotoma, by mobilizing and uploading new records, catching and fixing errors, adding new age models, and building new software or analytical workflows that directly link to Neotoma data. Examples of these two-way interactions include the HOPE project, which is helping facilitate uploads of data into the EPD and LAPD, and Wang et al. (2019), which has just built over 550 new Bacon age models for North American pollen records. Several papers describe Neotoma itself: 1) Williams et al. (2018) offers an overview of Neotoma, its design philosophies, and key concepts; 2) Grimm et al. (2018) reviews the concepts of constituent databases and data stewards; and 3) Goring et al. (2018) reviews the educational resources associated with Neotoma. And, the PAGES Fall 2018 issue on Building and Harnessing Open Paleodata (<https://doi.org/10.22498/pages.26.2>) provides a survey of community curated and open data resources across the paleosciences.

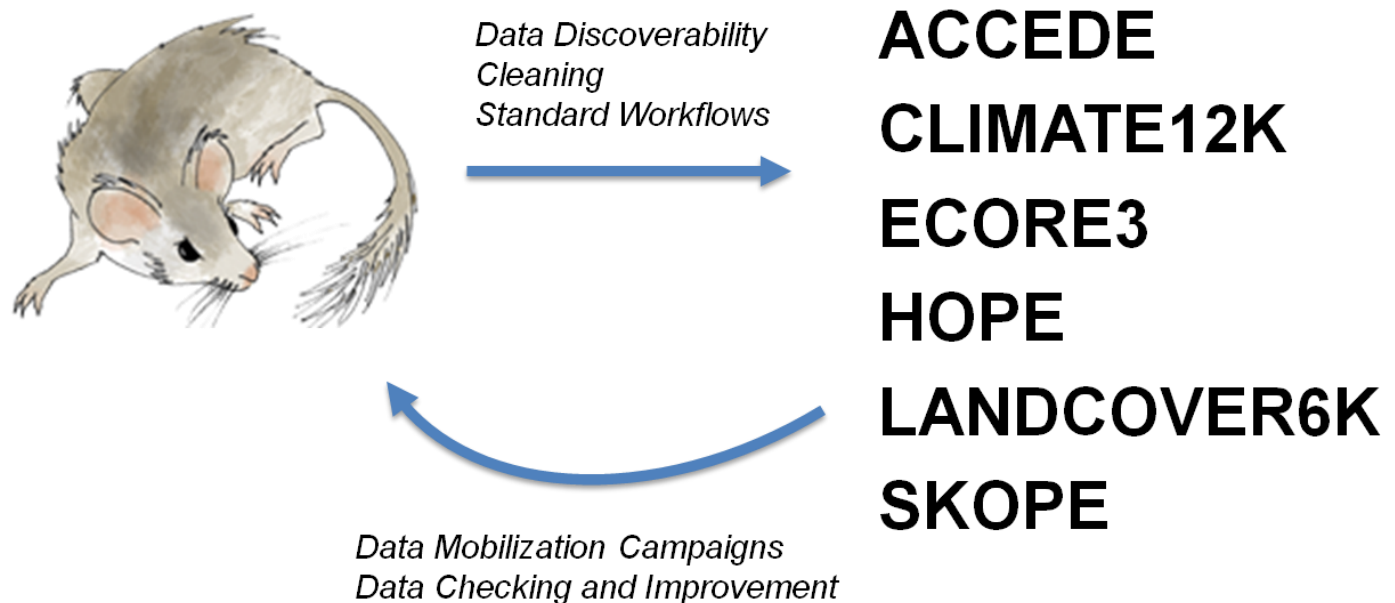


Figure: S.Goring

Data Services

1) DOIs. All Neotoma datasets are now assigned unique identifiers (DOIs), each associated with a landing page (Example 1: pollen dataset at Ballygisheen Bog: <http://data-dev.neotomadb.org/14194>; Example 2: vertebrate dataset at La Grande: <http://data-dev.neotomadb.org/5900>). These DOIs and landing pages support requirements for FAIR data publication by AGU and other journals (Wilkinson et al., 2016), and they can be readily found and searched by Google Datasets and other search engines. The automatic generation of DOIs is still in the beta stage and for now we generate DOIs in batches on an as-needed basis. We expect full automation by early 2020. 2) Flyover Country. Flyover Country (<https://flyovercountry.io/>) is a phone app designed for travelers interested in learning about the geological world around them. Travelers can enter their start and end locations, and then Flyover Country will provide them information about the geological formations that they see and fossil localities drawn from Neotoma and the Paleobiology Database.

Community

Neotoma supports and is supported by a large and growing community of data stewards, data contributors, and leaders. Particular shout-outs to Bob Booth (Testate Amoebae, Lehigh), Mona Dombosh (FAUNMAP, Penn State), Don Charles (Diatoms, Drexel), Suzette Flantua (LAPD, Bergen), Thomas Giesecke (EPD, Goettingen), Alison Smith (Ostracodes, Kent State), Alison Stegner (FAUNMAP, Stanford), and many more. And, most of all, Eric Grimm, for his tireless work and leadership as lead steward and Tilia engineer. The Leadership Council for 2019 is: Alan Ashworth, Jessica Blois (Assoc Chair), Phil Buckland, Thomas Giesecke, Simon Goring, Eric Grimm, Claudio Latorre, Suzie Pilaar-Birch, Alison Smith, Hiraku

Takahara, and Jack Williams (Chair). And thanks to Simon Goring (Wisconsin), Mike Stryker (Penn State), and Steve Crawford (Penn State) for keeping the lights on and new features rolling out.

If you're interested in learning more, see www.neotomadb.org and its resources page <https://www.neotomadb.org/about/category/resources>. If you'd like to become a member, you can sign up here: <https://tinyurl.com/NeotomaMember>. If you'd like to contribute data or become a steward, you can reach out to Jack Williams (jwilliams1@wisc.edu), Jessica Blois (jblois@ucmerced.edu), Eric Grimm (eric.c.grimm@outlook.com), anyone on the Leadership Council, or the all-purpose Neotoma email: neotoma-contact@googlegroups.com. Neotoma has been supported by NSF's Geoinformatics and EarthCube programs, and the Belmont Forum. Let us know if and how we can help you, your data, and your science.

References

- Goring, S.J., Graham, R., Loeffler, S., Myrbo, A., Ormand, C.J., Oliver, J.S., Williams, J.W. (2018) The Neotoma Paleocology Database: A research-outreach nexus. *Paleobiology Short Course*.
- Grimm, E.C., Blois, J.L., Giesecke, T., Graham, R., Smith, A.J., Williams, J.W. (2018) Constituent databases and data stewards in the Neotoma Paleocology Database: History, growth, and new directions. *PAGES Magazine* 26, 64-65.
- Wang, Y., Goring, S.J., McGuire, J.L. (2019) Bayesian ages for pollen records since the last glaciation in North America. *Scientific Data* 6, 176.



www.neotomadb.org/

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016) The FAIR Guiding Prin-

ciples for scientific data management and stewardship. *Scientific Data* 3, 160018.

Williams, J.W., Grimm, E.C., Blois, J., Charles, D.F., Davis, E., Goring, S.J., Graham, R., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A.C., Betancourt, J.L., Bills, B.W., Booth, R.K., Buckland, P., Curry, B., Giesecke, T., Hausmann, S., Jackson, S.T., Latorre, C., Nichols, J., Purdum, T., Roth, R.E., Stryker, M., Takahara, H. (2018). The Neotoma Paleoecology Database: A multi-proxy, international community-curated data resource. *Quaternary Research* 89, 156-177.

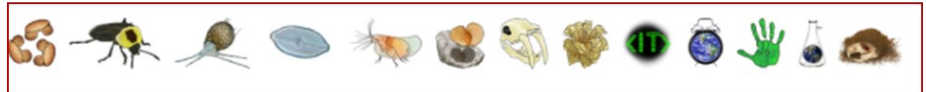
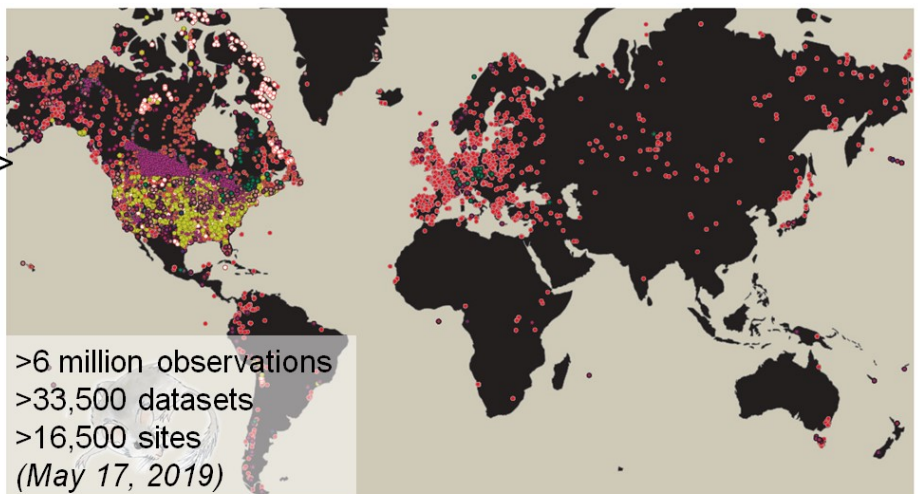
Neotoma Paleoecology Database



Mission: Support global-change research by providing an open, community-curated repository for multiple kinds of paleoecological data

Key Characteristics

- Open Data
- Curated by Community -> High-Value Data
- Standardized Variable Names & Taxonomy
- Flexible data models for sites, proxies, cores, etc.
- Time: Age Controls and Age Models



Fossil Pollen, Vertebrates, Diatoms, Ostracodes, Testate Amoebae, Stable Isotopes, Plant Macrofossils, Packrat Middens, ...

Williams et al. (2018) *Quaternary Research*