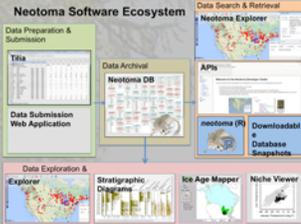


NEOTOMA PALEOECOLOGY DATABASE



Introduction

The Neotoma Paleocology Database (www.neotomadb.org) is a multi-proxy, open-access, relational database that includes fossil data for the past 3 million years from the late Neogene and Quaternary Periods. Modern distributional data for various organisms are also being made available for calibration and paleoecological analyses. The project is an international collaborative effort among individuals from more than 20 institutions worldwide, including domain scientists representing a spectrum of Pliocene-Quaternary fossil data types, as well as experts in information technology. Working groups are active for diatoms, insects, ostracodes, vertebrates, rodent middens, pollen/stratigraphy, age models, geochemistry and taphonomy. Groups are also active in developing online tools for data analysis and for developing modules for teaching at different levels.



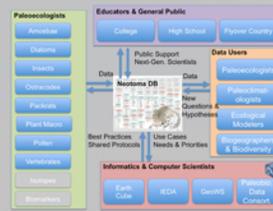
Design Principles

The design and implementation of the Neotoma Paleocology Database is informed by several fundamental design principles:

- Community-led Distributed Scientific Governance.** Distributed scientific governance is essential for Neotoma because no single individual or institution can be expert in all the taxonomic groups, regions, and time periods stored within Neotoma. Data Stewards from various domains and regions coordinate, manage, and vet data before uploading to the database.
- Consolidated IT.** Neotoma provides shared cyber-infrastructure that reduces developer support costs, enhances efficiencies, maximizes data interoperability, and scales as new records and proxy types are added.
- Open Data.** Neotoma uses a CC-BY license, allowing free reuse of data with proper attribution to Neotoma and the original data contributors. By supporting open data, Neotoma helps promote scientific transparency and reproducibility.
- Neotoma is a Living Database.** The life cycle of a paleoecological dataset does not end with its publication. Data are re-used, errors are corrected, and derived datasets updated with newer analytical methods. Neotoma retains both the original and the changing dimensions of these data, facilitating both reproducibility and ongoing discovery.
- Low Data Friction.** Our goal is to make it as easy as possible to upload data to and retrieve data from Neotoma, via an ecosystem of software tools. Some tools facilitate data entry and quality-checking, while others aid data discovery, exploration, and retrieval.
- Community Engagement and Empowerment.** Neotoma serves a diverse community. These groups are essential to scaling Neotoma, by engaging and empowering a community of Data Stewards, contributors, and third-party software developers. The sustainability of Neotoma is ultimately determined by its ability to undertake the key research priorities of these users.

Education & Outreach

In collaboration with the Academy of Natural Sciences of Drexel University (ANS), a new series of standards-based climate science lessons for high school students will be created. Neotoma fossil data, with the proper scaffolding, can provide an access point that teachers and students can use to investigate earth system changes, develop an understanding of past biotic responses to climate change, and apply knowledge gained to current rates of environmental change. ANS will also pilot the lessons with its Women in the Natural Sciences Program (WINS), a free after-school and summer science enrichment program for high school girls. Since its founding in 1982, WINS has introduced hundreds of high school girls to future careers in science and other professions by providing hands-on science workshops, career and college exploration, and positive youth development. Feedback from WINS teachers and students will be used to construct the final set of lessons for production and publication. Lessons will be available on the Neotoma website and the ANS website.



Authors

- Allan Ashworth, Dept. of Geosciences, North Dakota State University
- Julio Betancourt, National Research Program, U.S. Geological Survey
- Brian Bills, Center for Environmental Informatics, Penn State University
- Jessica Blois, School of Natural Sciences, University of California, Merced
- Bob Booth, Dept. of Earth and Environmental Science, Lehigh University
- Phil Buckland, Dept. of Historical, Philosophical and Religious Studies, Umeå University
- Don Charles, Dept. of Biodiversity, Earth and Environmental Science, Drexel University
- Eric Grimm, Dept. of Botany, Illinois State Museum
- Russell Graham, Dept. of Geosciences, Penn State University
- Sonja Hassmann, Academy of Natural Sciences, Drexel University
- Alison Smith, Dept. of Geology, Kent State University
- Jack W. Hillam, Dept. of Geography, U. of W. Izonah

Insects

The archival, analytical, and educational goals for the Insects working group, headed up by Allan Ashworth and Phil Buckland, for the next cycle include:

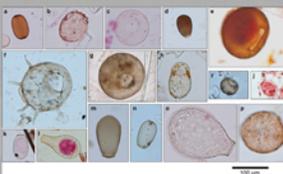
- Complete uploading of fossil Coleoptera from 120 North American sites
- Add North American Chironomidae to the database
- Improve the online display of data -- will require editing of existing data
- Ingest and mirror the Bugs, Coleopteran Ecology Package (www.bugsecp.com), the world's largest database of Quaternary fossil insect records. It will also connect to the Strategic Environmental Archaeology Database (www.sead.ac.uk), enhancing access to a wide variety of European environmental archaeology datasets. This cooperation will increase the accessibility and ultimately data interoperability and reuse potential of a large number of datasets containing proxy data on past environments, people and activities.
- Add data from Russia, Japan, Australia, New Zealand and South America.
- Add modern distributional, ecological and climatic data for species in the data base
- Develop paleoecological and paleoclimatic analytical tools
- Develop a teaching module so that students will be able to make their own paleoclimatic analyses



Testate Amoebae

Testate amoeba data from lacustrine and peatland environments have increased dramatically in the past decade, yet these are not currently archived in any systematic way despite strong community interest. These data include surface calibration datasets as well as paleoecological records and derived values like reconstructed water-table depths. Over the next few years, the testate amoeba working group will:

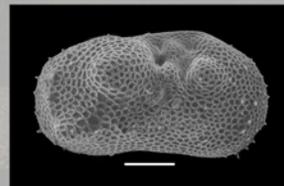
- Build a taxonomic hierarchy for the database
- Upload surface-sample calibration datasets and fossil datasets
- Train data users and data stewards. Robert Booth is leading this effort and will be seeking feedback directly from the testate amoeba research community
- Plan workshops on the database and its development for the 2016 and 2018 International Society of Testate Amoeba Research (ISTA) meetings. Robert Booth is leading this effort and will be seeking feedback directly from the testate amoeba research community



Ostracodes

The Neotoma Ostracode Working Group, headed up by Alison Smith and Brandon Curry, has two immediate goals:

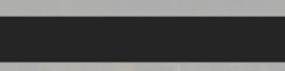
- We plan on continuing the upload of modern samples and their associated water chemistry and climate data from NANODE and NAACODE (Canada and U.S.A.) data.
- Also, we plan on uploading Plio-Pleistocene records of fossil ostracodes from key paleolimnologic sediment cores taken in western North America.
- By uploading these data into Neotoma, we will be able to search and map species distributions, and build maps that compare the ostracode distributions with distributions of other fauna and flora from Pliocene to Present.
- A goal for the future is to expand the modern ostracode database in Neotoma to include databases from outside North America, such as NODE (Nonmarine Ostracode Distribution in Europe). In particular, we are very focused on achieving a Holarctic, circumglobal mapping view, but we continue to be interested in data from around the globe.



Diatoms

The main priority of the Diatom working group, headed up by Don Charles and Sonja Hassmann, over the next four years will be adding more datasets to Neotoma.

- There are over 200 datasets in our working list of potential useful datasets to add, and many commitments by contributors. Priority will be given to contributors with datasets ready to enter, and that will be most useful for providing wide ranging geographic coverage, adding value to existing datasets, and most relevant to key environmental issues such as climate change.
- Additions will include diatom count data to core stratigraphies and calibration datasets, chronologies inferred environmental values, and related metadata.
- Considerable effort will be devoted to adding to and maintaining the diatom taxonomy list. In addition, we propose to fully develop the synonym list so that it will be much easier to directly compare results among multiple datasets.
- We will also hold workshops and training sessions to show contributors how to format their data for entry into Tilia. Some of these may be in conjunction with diatom meetings such as the North American Diatom Symposiums in fall 2013 and 2017.



Middens

The Neotoma Packrat Midden Working Group's near-term goals are:

- Add new midden data from North and South America. Claudio Latorre is processing rodent-midden data from South America, and has uploaded several datasets.
- Relate ancillary analyses (e.g., pollen, vertebrates, isotopes, ancient DNA, morphological measurements) made after initial publication of the plant macrofossil series.
- Interface the existing North American Packrat Midden Database for Plant Macrofossils (NAMPD) into Neotoma. Julio Betancourt and Kate Rylander have obtained complete fossil lists from the original authors for ~180 of these studies and are entering these data into Tilia; complete data are entered for ~130 sites, and ~50 more are partially entered. In addition, Steve Jackson has contributed data from ~400 middens from 35 sites in Wyoming. After these complete datasets are uploaded, additional sites from the USGS database will be processed and uploaded.
- Use Neotoma to conduct gap analyses that motivate meta-analysis of midden records in the context of other late Quaternary information, guide future field campaigns, and inspire novel applications of midden data and materials.



Pollen & Plant Macros

New pollen datasets are being generated at a high rate, and legacy data are still being acquired.

- This working group, headed by Eric Grimm, will continue to upload data for various pollen database projects, including the North American Pollen Database, Latin American Pollen Database, European Pollen Database (EPD), Indo-Pacific Pollen Database, and Japanese Pollen Database.
- Data from the African Pollen Database are slated for inclusion in Neotoma.
- The PAGES LandCover 6k project launched in 2011 intends to use Neotoma as its primary database and has a stewards training workshop planned for June 2016.
- Data Stewards for pollen will process plant macrofossil data, and plant macrofossils associated with new pollen datasets usually are uploaded simultaneously. Directly dated and identified plant macrofossils in 14C geochemistry datasets are automatically entered into Neotoma as plant macrofossil datasets.
- The legacy data from the North American Plant Macrofossil database are slated for inclusion into Neotoma.



Vertebrates



Building on the success of Phase One of the Neotoma database development, over the next few years the Vertebrate Working Group, headed by Jessica Blois, Russ Graham, Ed Davis, will focus on:

- Improving the scope of the base data within Neotoma by entering new datasets. We will incorporate new databases into Neotoma (e.g., Mammap, the ANTIQUA database focused on South American megafaunal extinctions, and MQMD, the Mexican Quaternary Mammals Database). We will also start moving recently validated FAUNMAP 2 data into Neotoma.
- Improving the quality of existing data in Neotoma. We will revisit the data entered for localities, correct any mistakes, and add new chronological information or associated taxa, if available.
- Enhancing ability to store specimen-level data. Storing specimen-level data will facilitate linkages to museum collections, GenBank, radiocarbon dates, taphonomic information, and many other kinds of associated data.
- Enabling distributed data entry and use. Along with other working groups, we will continue to facilitate data entry by, e.g., developing easy-to-use scripts that can automate data entry, training additional data stewards, developing data analysis examples in R, and developing outreach and training workshops for new users.

Development of Analytical Tools

The Analytical Tools working group, headed by Jack Williams and Simon Goring, are developing the following applications:

- Ice Age Mapper** will provide an interactive, animated mapping system for visualizing the dynamic patterns of species distributions in North America over the late Quaternary, designed to show users how species ranges shifted in response to the climate changes and melting ice sheets during the last deglaciation.
- Niche Viewer** will support the flexible plotting of taxa in environmental space using surface sample datasets to build paleoclimatic transfer functions that predict environmental variables in the past as a function of species presence or abundance across space today, enabling quantitative inferences about past ecosystems and environments. Proxy data types to be incorporated into Niche Viewer include pollen, insects, ostracodes, diatoms, and testate amoebae.



Authors

