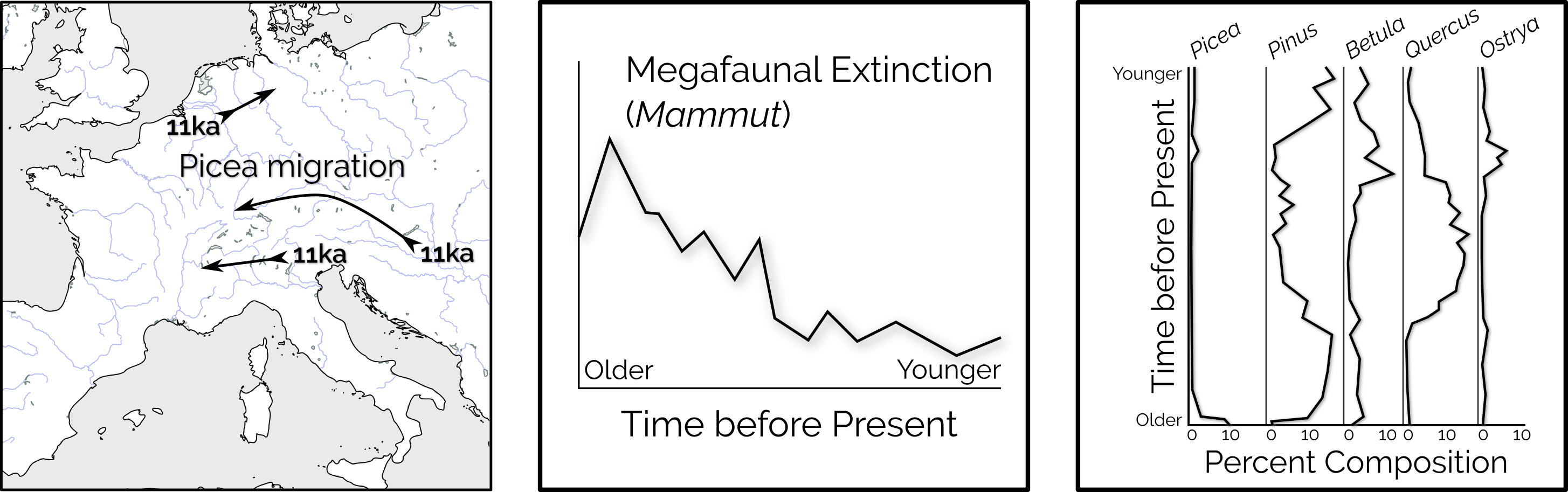
# Neotoma Paleoecology Database: A research-outreach nexus

## Introduction

The Quaternary period is well suited to a number of educational outreach opportunities. The last 2.5 million years represent a geological time scale that overlaps with the evolution, migration, and eventual development of human societies. Thus, the defining features of the Quaternary are both proximate in time, and highly relatable for many. In addition to the development of much of what we recognize as modern human culture, the Holocene (the last 11,700 years) represents a time during which the peopling of North America and its subsequent colonization occurred, the last continental ice sheets collapsed (excluding Antarctica), sea levels rose, and charismatic megafauna went extinct. The Quaternary is a geological period that intersects with geophysical change at human time scales, and it is a period that has been intensely studied with an eye to using it as a proxy for future climatic change.



**Figure 1**. Things you can study in the Quaternary: (1) Species migration through time, here, *Picea* migration after Latalowa and van der Knaap (2006); (2) Megafaunal extinction (after Gill et al. 2009) and (3) forest community change through time represented in pollen archives (after McAndrews 1966).

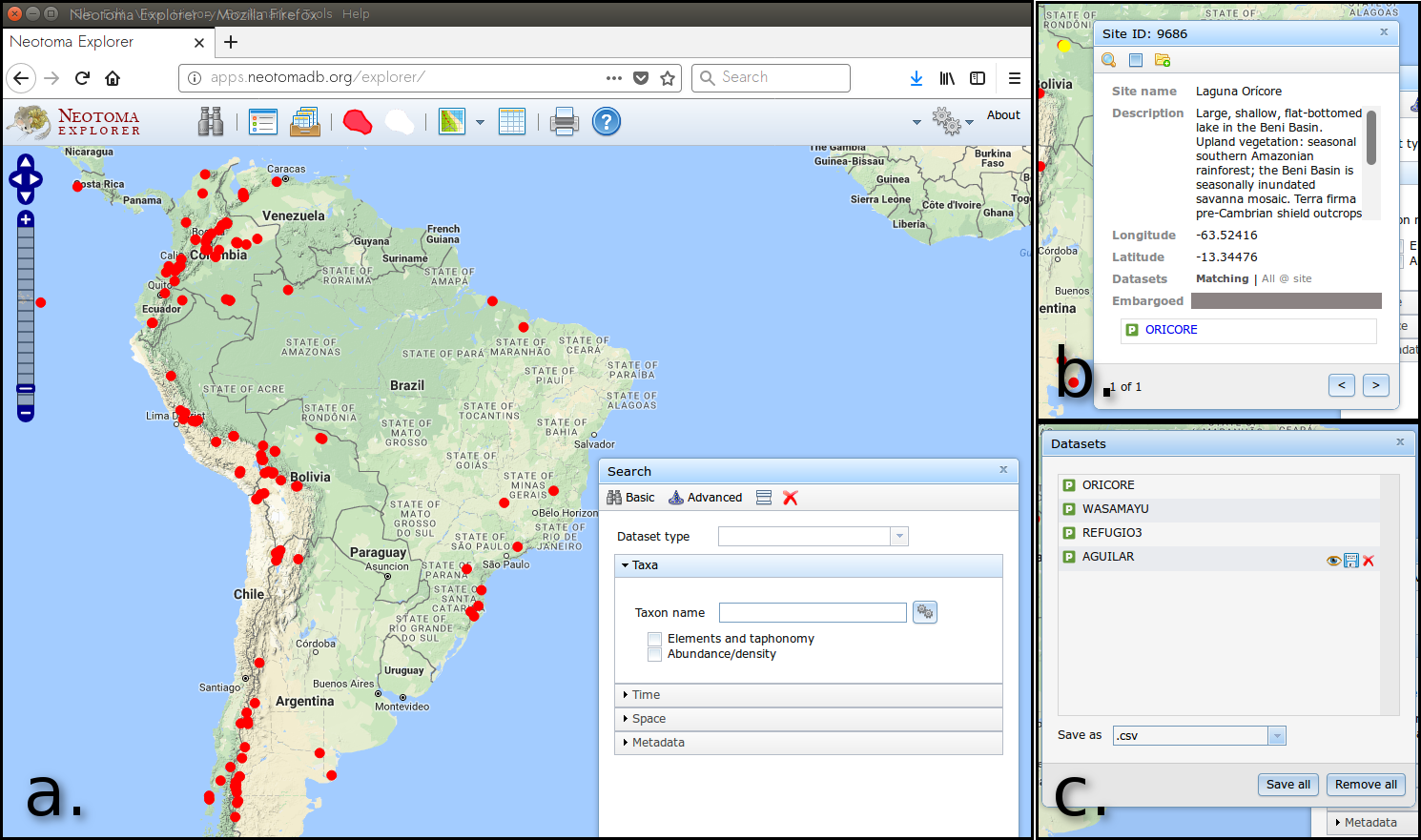
The study of past ecosystems, whether from excavations, lake sediment sampling, or other fossil localities, results in a complex set of data, including information about age, taxonomy, spatial coordinates, depth or stratigraphy, and other related information. This information can be difficult to understand, but the Neotoma Paleoecological Database provides a public portal for data access that can be used to understand ecological change through the Quaternary, and recent outreach activities have resulted in the production of a number of curriculum tools focused on upper secondary and post-secondary students.

This chapter introduces the Neotoma Paleoecological Database as a tool for outreach, framed as an educational activity. We explore several other educational outreach activities, including outreach activities centered on upper level undergraduate and graduate students using the software package R. We conclude this chapter by examining the role of community curated data resources as bridges between the research community and the public, with regard to educational outreach in particular.

## An Introduction to Neotoma

Neotoma is community curated data resource that contains more than XXXX unique observations of fossil organisms and associated geophysical information spanning the last 2.6 million years. These records and their associated metadata are contained within a web-accessible database that has been designed for research and outreach activities (J. W. Williams et al. in press).

Much work done by paleoecologists uses a single site, or several sites from which proxy data are obtained, and for which ecological inference is derived. The Neotoma Paleoecological Database (http://neotomadb.org) was the result of the paleoecological community coming together to engage in a data-model intercomparison with early general circulation modelers (J. W. Williams et al. in press; P. Anderson et al. 1988). From this initial stage, Neotoma has successfully engaged with other research communities and evolved to become a critical element within the infrastructure of the paleoecological research community.

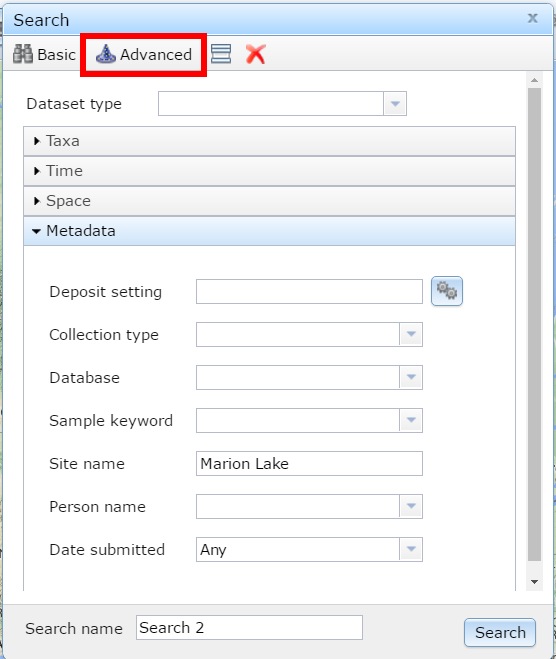


**Figure 2**. *The Neotoma Explorer is a web based data discovery tool, that allows a map based search for paleoecological sites using species names, site locations, site names, or other associated metadata (a). The tool provides the opportunity to examine individual sites (b) and to select a number of sites for download and later processing or study (c).*

Neotoma uses a specific, but flexible data model to describe paleoecological data records (J. W. Williams et al. in press; E. Grimm 2008). The Neotoma data model contains information about researchers, locations, publications, time and the raw paleoecological records that underpin our understanding of Quaternary environments. Neotoma is both a repository and a data access portal. The primary data discovery interface for Neotoma is the Neotoma Explorer (Figure 2), powered by an Application Program Interface (API: http://api.neotomadb.org) that can be directly accessed through the world-wide web. The use of an API provides a mechanism to build new tools on top of Neotoma's data resources. The API has allowed the development of the neotoma R package (S. Goring et al. 2015) and the inclusion of Neotoma data within the Flyover Country app (http://fc.umn.edu/).

The following exercise will help individuals understand how paleoecologists interpret and process proxy data, from a single site, to continental scales, from the near-modern to deep time records. It will teach users how to search for individual paleoecological sites, find publications and visualize change through time. It will also introduce the kinds of research questions, for example, species co-occurrence, that paleoecologists are interested in.

### Your First Search



**Figure 3**. *Searching for Marion Lake using the Neotoma Explorer's search tools.*

The [Neotoma Explorer](apps.neotomadb.org/explorer) provides the spatial overview and search capabilities for Neotoma. One of the simplest things you can do with the database is to search for a single research site. To do this, look for the "*Metadata*" search option under the "*Advanced*" tab of the search panel (Figure 3). "Marion Lake" is an important paleoecological record from western North America, the record provided one of the first quantitative reconstructions of Holocene climate from western North America (Mathewes and Heusser 1981). *Search for Marion Lake*. You may find more than one site. Which one is the pollen dataset? How do you know?

The northern-most site is the Marion Lake pollen sample site. When you click on the point you should see some site information and a description, along with a green **P**, as well as a small clock icon. Both have the word MARION beside them. The icons represent pollen data and geochronological data respectively. *Click on the pollen data* first and scroll through the tabs available to you. You'll see:

* **Samples** -- Individual pollen counts, depth information and summary chronology.
* **Diagram** -- For data where diagrams are possible (pollen, diatoms, etc.) a tool to draw the change in the data with time.
* **Site** -- Site level information for the data.
* **Chronology** -- The age model used for the record.
* **Publications** -- Publications related to the dataset.

These tabs provide you with a pretty good overview of the dataset, long term changes at the site, and any papers you might want to read to get more information about the record.

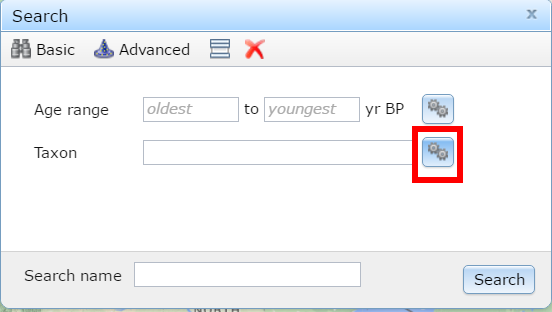
#### More Searching

Neotoma can also be used to look for a large number of datasets with a particular dataset type. *Create a new search for a* ***Dataset Type*** *(make sure you clear the site name)*. What do you see? How does that dataset look different from the pollen dataset we looked at earlier? Why?

#### More Details

Now *click on the white polygon on the top bar*. If you enter a date (say 21100), hit enter, and then zoom out. At its full extent, you'll see a white polygon, representing the continental ice sheets, overlain on the map of sites. It is possible to use the ice sheet distributions to help understand paleo-distributions of taxa, their presence or absence.

#### Research Questions



Mastodon is a progressive heavy metal band, thrice nominated for a Grammy award, but it was also a species of extinct mammutid proboscideans.

We tend to picture late-glacial environments as being cold, similar to tundra environments. Places like Beringia spring to mind. *Do a taxon search for* Mammut *(mastodons) and look at the site locations that appear*. What happened?

More records can be found by doing a multi-taxon search. *Click the gear box to search for multiple taxa*. Search and select all the *Mammut*s we can find in the list. Make sure to choose the right *Taxa Group* and then search. Where are the Mastodons?

We can also look for sites with common forest taxa. Let's look for sites with *Picea* in them, and, since *Mammut* went extinct somewhere around 12,000 ybp (Gill et al. 2009), lets limit our search to all sites with *Picea* taxa, abundance over 20% and a time range of 15,000 -- 12,000 ybp. What can we say about our friends the mastodons?

### Summarizing The Lesson

We've learned about Neotoma now.

## The Role for CCDRs in Outreach

The development of community supported data resources is often a by-product of a community of researchers attempting to answer research questions that can only be answered through the use of a number of datasets with some standard data representation. Traditionally, Community Curated Data Resources (CCDRs) have been a resource for data storage: systems of record; collating, aggregating and cleaning data, largely supervised by domain experts (Kapoor et al. 2015). As mechanisms to access and visualize data have improved across the geosciences, it has become possible to generate opportunities for individuals and groups outside of the traditional research community to access and interact with this data, moving these databases from systems of record to systems of *engagement*, adding a social and collaborative element to the platform. As such, engagement with educational activities represents a major opportunity to help transform CCDRs, providing an opportunity for broader engagement, and for the development of a robust user community, adding value beyond just data content.



**Figure 1**. CCDRs involved in engagement can access new communities outside of the traditional sphere of activity in the research community.

By acting as a portal for outreach with the research community, CCDRs serve an important role in engaging the public, and providing a service to researchers by centralize some outreach activities, strengthening links to disciplinary experts in the science education community. For example, the NSF Funded project "Neotoma: Community-led Cyberinfrastructure for Global Change Research" was able to leverage a broad group of researchers, and connect with Earth Science Educators in the Science Education Resource Center (SERC) at Carleton College (https://serc.carleton.edu/index.html) to generate a range of high-quality educational resources. Successful educational outreach benefits strongly from engagement with skilled educators. Earth Science education, and pedagogical research have matured significantly over the past years. Thus, CCDRs provide an opportunity to link research communities to educators, and educators to the broad range of activities undertaken within the communities, building on the strengths of each group.

The emerging field of paleoecoinformatics (Brewer, Jackson, and Williams 2012) has enhanced the paleoecology communities suite of tools, bringing geospatial tools, and the technical capacity to combine web or mobile mapping tools with data resources. The development of the Neotoma Explorer (http://apps.neotomadb.org/explorer) provides an easy interface for elementary and secondary school students, while access through the Neotoma API and the neotoma R package (S. Goring et al. 2015) allows the integration of paleoeological data into post-secondary curriculum.

## Outreach Venues

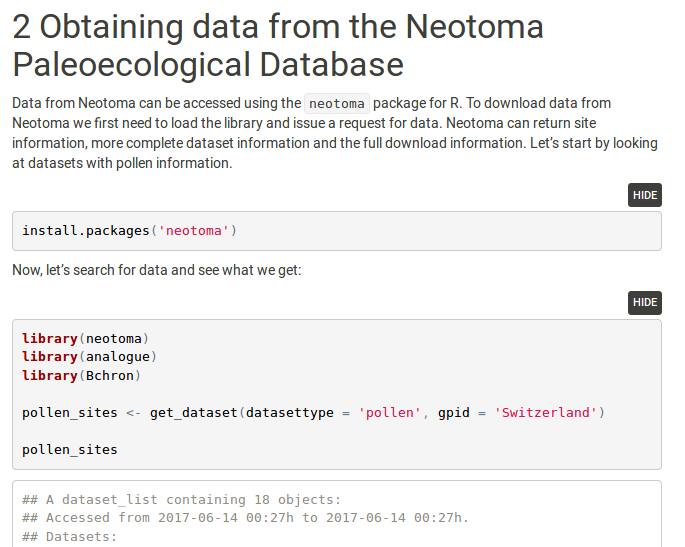
We highlight two particular educational outreach opportunities. The first, as above, are associated with the 2015 SERC Workshop ["Using the Neotoma Database in the Classroom"](https://serc.carleton.edu/neotoma/workshop/index.html). The SERC materials are largely The second are a set of curated resources associated with research activities.

### SERC Workshops

The 2015 Carleton College Neotoma workshop resulted in the development of several workshop resources. The workshop provided an opportunity to train paleoecology domain experts in educational outreach, and to develop high-quality teaching materials at various educational levels. This paper focuses on three teaching units, focused on undergraduate, or senior high school students: (1) ["Exploring the Neotoma Paleoecology Database"](https://serc.carleton.edu/neotoma/activities/121251.html), material directed toward college undergraduates that covers mammal and vegetation distribution changes since the last Glacial Maximum; (2) ["Species distributions in response to environmental gradients in the Upper Midwest of the United States - an example using the Neotoma database"](https://serc.carleton.edu/neotoma/activities/121269.html), a unit focused on environmental gradients over short spatial scales and how these change over time; and, (3) ["Climate Change and Mammal Dispersal"](https://serc.carleton.edu/neotoma/activities/121239.html) which is directed toward senior high school and early undergraduate students, helping them understand dispersal in small mammals through time in response to climate change.

### Programmatic Access and Paleoecology

Neotoma supports data access through an Application Program Interface (API) that provides users with the ability to access data from Neotoma using programming languages such as JavaScript, R or Python. Since the publication of Goring *et al*. (S. Goring et al. 2015), a number of resources using the API have been published online in a publically accessible format. In particular, [GitHub](http://github.com) has the ability to host programmatic code, which acts as a resource, but also publically hosted pages (using [GitHub Pages](https://pages.github.com/)), which provides the opportunity to host educational material, along with the source code that powers any programmatic workflows.



**Figure 2**. Educational material presented on [Neotoma's GitHub Pages](https://neotomadb.github.io/workbooks/AgeModels.html) show the initial stages of constructing chronologies for paleoecological records.

#### Chronology Construction & CLimate

Chronology construction and pollen-based climate reconstruction are both complex tasks. Chronology construction is central to the field of the paleogeosciences (M. Harrison et al. 2016), while climate reconstruction is widely used, but fraught with methodological issues (Juggins 2013). Within the Quaternary geosciences there are a number of techniques that can be used to construct chronologies, and packages that can be used to reconstruct climate from proxy data. Tools such as Bacon (Blaauw and Christen 2011) and BChron (Parnell 2016) for chronologies, or rioja (Juggins 2017) and analogue (G. L. Simpson 2007) for climate reconstruction, have a number of settings and requirements, however, the paleoecological community is highly distributed, and many students might not encounter a course on chronology construction or paleoclimate reconstruction. In this way, the CCDR acts as a source of best-practices information, and is able to show how this chronology construction works within a workflow that includes the CCDR itself.

#### Workshop Materials

Neotoma hosts workshop materials on its GitHub page. These materials use the software principles of "version control" to borrow elements across workshops, while allowing workshop leaders to take the strengths of core workshop materials. This provides the ability to both reuse well established teaching materials and to tailor for new audiences. Workshop material for five separate workshops is currently available in the [Neotoma Workshops GitHub repository](https://github.com/NeotomaDB/Workshops), and new workshop materials are being actively developed.

### Beyond the Classroom

Flyover Country section. How technological developments within the CCDR community can lead to unexpected developments and opportunities.

## Conclusions

Community supported

## References

Anderson, PM, CW Barnosky, PJ Bartlein, PJ Behling, L Brubaker, EJ Cushing, J Dodson, et al. 1988. “Climatic Changes of the Last 18,000 Years: Observations and Model Simulations.” *Science* 241 (4869). American Association for the Advancement of Science: 1043–53.

Blaauw, Maarten, and J Andrés Christen. 2011. “Flexible Paleoclimate Age-Depth Models Using an Autoregressive Gamma Process.” *Bayesian Analysis* 6 (3). International Society for Bayesian Analysis: 457–74.

Brewer, Simon, Stephen T Jackson, and John W Williams. 2012. “Paleoecoinformatics: Applying Geohistorical Data to Ecological Questions.” *Trends in Ecology & Evolution* 27 (2). Elsevier: 104–12.

Gill, Jacquelyn L, John W Williams, Stephen T Jackson, Katherine B Lininger, and Guy S Robinson. 2009. “Pleistocene Megafaunal Collapse, Novel Plant Communities, and Enhanced Fire Regimes in North America.” *Science* 326 (5956). American Association for the Advancement of Science: 1100–1103.

Goring, Simon, Andria Dawson, Gavin Simpson, Karthik Ram, Russ Graham, Eric Grimm, and John Williams. 2015. “Neotoma: A Programmatic Interface to the Neotoma Paleoecological Database.” *Open Quaternary* 1 (1). Ubiquity Press.

Grimm, EC. 2008. “Neotoma: An Ecosystem Database for the Pliocene, Pleistocene, and Holocene.” *Illinois State Museum Scientific Papers E Series* 1.

Harrison, Mark, Suzanne Baldwin, Marc Caffee, George Gehrels, Blair Schoene, David Shuster, and Brad Singer. 2016. “Geochronology: It’s About Time.” *Eos* 97 (5). American Geophysical Union: 12–13.

Juggins, Steve. 2013. “Quantitative Reconstructions in Palaeolimnology: New Paradigm or Sick Science?” *Quaternary Science Reviews* 64. Elsevier: 20–32.

———. 2017. *Rioja: Analysis of Quaternary Science Data*. <http://www.staff.ncl.ac.uk/stephen.juggins/>.

Kapoor, Shubir, Aleksandra Mojsilovic, Jade Nguyen Strattner, and Kush R Varshney. 2015. “From Open Data Ecosystems to Systems of Innovation: A Journey to Realize the Promise of Open Data.” In *Bloomberg Data for Good Exchange Conference*. <https://pdfs.semanticscholar.org/ebfa/cdfc9da14c5b54791e6fba89bd7a6c7809d0.pdf>.

Latałowa, Małgorzata, and Willem Oscar van der Knaap. 2006. “Late Quaternary Expansion of Norway Spruce Picea Abies (L.) Karst. in Europe According to Pollen Data.” *Quaternary Science Reviews* 25 (21). Elsevier: 2780–2805.

Mathewes, Rolf W, and Linda E Heusser. 1981. “A 12 000 Year Palynological Record of Temperature and Precipitation Trends in Southwestern British Columbia.” *Canadian Journal of Botany* 59 (5). NRC Research Press: 707–10.

McAndrews, John H. 1966. “Postglacial History of Prairie, Savanna, and Forest in Northwestern Minnesota.” *Memoirs of the Torrey Botanical Club* 22 (2). JSTOR: 1–72.

Parnell, Andrew. 2016. “Package ‘Bchron’.” <https://cran.r-project.org/web/packages/Bchron/index.html>.

Simpson, Gavin L. 2007. “Analogue Methods in Palaeoecology: Using the Analogue Package.” *Journal of Statistical Software* 22: 1–29.

Williams, John W., Eric C. Grimm, Jessica Blois, Donald F. Charles, Edward Davis, Simon J. Goring, Russell W. Graham, et al. in press. “The Neotoma Paleoecology Database: A Multi-Proxy, International Community-Curated Data Resource.” *Quaternary Research*.