

Lab 1: Exploring Paleoecological Data with Neotoma Explorer

Lesson Overview

This lab is designed to introduce you to the Neotoma Paleoecology Database and Neotoma Explorer. Neotoma DB is an open and community-curated resource for paleoecological data. Open means that anyone can find and download data from Neotoma. Community-curated means that Neotoma data are augmented and maintained by a community of experts, with many scientists contributing their data to Neotoma and actively working to curate and improve these data. Explorer is a graphical interface that is very handy for a first-pass search, discovery, and viewing of paleoecological data, typically exploring a few sites at a time. In following weeks, we will move to more advanced visualizations and analyses of paleoecological data, using software designed to download data from Neotoma and analyze it.

Lesson Goals

1. **Gain familiarity with Neotoma Explorer interface**, as tool for discovering and exploring paleoecological data
2. Learn how to **search for data**, using various search options.
3. Learn how to **make quick maps** of past species distributions over time
4. Learn how to **view** data, using stratigraphic pollen diagrams.
5. Learn how to **download** paleoecological datasets from Neotoma.
6. Do a bit of thinking along the way about **ecological patterns and processes** revealed by these first-pass analyses.

This lab is intentionally somewhat light on instructions. Neotoma Explorer is designed to be a fairly intuitive interface for searching, viewing, and downloading data. Hence, these instructions are designed to point you in the right direction but leave some steps for you to figure out. If you run into questions, please reach out to me by Slack, so that others can easily see the Q&A.

For all **QUESTIONS** – write down the answer and turn this in as part of your assignment. For this exercise, most of these Questions are pretty simple – the main intent is to make sure that you are getting familiarity with Neotoma Explorer as a portal for discovering and quick-look visualizations of data. A few questions ask you to think and reflect about ecological patterns and processes revealed by this analysis. Please turn in your assignment using Canvas/Assignments.

Deadline: before next lab.

Note: Some people prefer to work solo; others like to work in small teams. Either is fine! If you choose to work in a small team (recommended maximum of 2 people), be sure to take turns. Typically, the student actively working at the computer learns much more than the person passively watching. If you choose to work as a team, you can turn in just one homework (with both your names on it) – but please turn in two copies of your work – one for each of you. That helps me track overall turn-in rates. Please use the following filename convention: **G523_F20_Lab1_[LASTNAME].[EXT]** For example:

G523_F20_Lab1_WILLIAMS.docx

I. Getting Started

1. Go to <http://www.neotomadb.org/>
2. Click on the 'Explorer' picture
3. Change the zoom so that your window is centered on North America, including all of the lower 48 states and the southern half of Canada.

II. Search for Data

1. Find a Site

- a. Goal: Find a site of interest
- b. Open the Search Window. This should automatically open when you launch Explorer, but if not, click on the binoculars icon.
- c. Using the Search Window dialogue (Metadata tab, Site Name field), find this site: Devil's Lake, WI.
 - i. **Tip**: Try searching for just 'Devil' – Neotoma will return all sites with this word in the site name.
 - ii. To search, press the 'search' button in lower right of the Search Window.
 1. **Tip** – the 'search' button sometimes will be hidden just offscreen below the metadata fields
- d. Once you've found it, click on the dot. A window will pop up with some information about this record.
- e. **QUESTION 1**: What is the latitude and longitude of Devil's Lake? What is the Site ID? [Note: Apparently the Site ID assignment for Devil's Lake is just a coincidence...]

2. Find a Taxon

- a. Goal: Find all sites with at least 20% *Picea* (spruce) pollen, between 21,000 and 18,000 years ago. (This time period corresponds to the last glacial maximum (LGM), when the Laurentide and Cordilleran ice sheets were at their maximal extents.)
- b. Search Window:
 - i. **Tip**: Before starting a new search, be sure to clear previous search fields! Click the red X to clear prior search fields.
 - ii. Dataset type: Set to 'pollen'
 - iii. Taxa: Set name to 'Picea', click the 'Abundance/density' field, then set to >20%
 - iv. Time: Set the age range from 21,000 (oldest) to 18,000 (youngest)
 - v. Space: in the Extent box, set to 'current map extent'
- c. **QUESTION 2A**: What state east of the Rockies has the southernmost location of *Picea* at the LGM?
- d. Run a second search for sites with at least 20% *Picea* (spruce) pollen, between 1000 and 0 years ago.
- e. **QUESTION 2B**: What state east of the Rockies has the southernmost location of *Picea* at this time? (Note that in reality, there are a few scattered southern outliers of *Picea* populations at the top of the southern Appalachian Mountains. Be careful about interpreting actual species ranges from occurrence data of variable density!).

3. Find all sites produced by a researcher

- a. Search Window: Metadata tab, Person Name field)
- b. Find all sites produced by Louis (Lou) Maher (a former palynologist in UW-Madison Geosciences).
- c. **QUESTION 3:** In which states did Lou work?

4. Find all fossil pollen records in North America

- a. Search window: Advanced Tab, Dataset field, click on down arrow at right to get a dropdown list of all dataset types.
 - i. For this exercise, choose 'pollen', not 'pollen surface sample'
 - ii. Set your map extent to be North America (as close as you can) and under 'Space' set Location to Search By Extent / Current Map Extent
- b. **QUESTION 4A:** How many sites and datasets did you get? (Hint: Watch for the yellow box in upper right of screen during the search return. If you miss it, you can try again.
- c. **QUESTION 4B:** The generation of fossil pollen records is labor intensive and hence expensive – e.g. the costs of fieldwork, the money spent on radiocarbon dates, the time required for a trained analyst to identify and count pollen data, etc. A very rough time/cost estimate for a single fossil pollen record is on the order of two years and \$30,000. Given this, give an order-of-magnitude estimate of the number of person-years and dollars it took to generate these fossil pollen records now stored in Neotoma. (Order-of-magnitude = 10 person-years? 100 person-years? 1000? etc.)
- d. **QUESTION 5:** Notice there is a fairly dense network of fossil pollen sites in an arc stretching from Minnesota to Nova Scotia. Why might there be such a dense network in this region?

5. Show/Hide Search Results

- a. You now have many search layers – it may be getting a bit confusing. Find the icon that lets you show/hide/combine search layers and use it to hide some of your searches. The icon is in the top gray bar and is next to the binoculars.

6. Multi-Taxon Search

- a. Find all sites with at least 20% *Picea* pollen between 15,000 and 12,000 years ago.
- b. Find all sites with *Mammut* (mastodon) between 15,000 and 12,000 years ago.
 - i. For *Mammut*, use the 'Advanced Taxon Selection' which you can use by clicking on the gears icon to the right of the 'Taxon' field in the 'Search' window.
 1. In the ATS window, choose 'Mammals' for Taxa Group (click on the down arrow to get a dropdown list of options) and then enter *Mammut* into the 'Search for' window. Then click 'Go'
 2. Note that the search returned taxon names for both *Mammut* (mastodon) and *Mammuthus* (mammoth). Click all boxes for all variants of *Mammut* but do not click the *Mammuthus* boxes
 3. Enter a search name (e.g. '*Mammut* – all') and click Save
 4. In the general search window, click 'Search'

- c. **QUESTION 6:** Take a screenshot of this map and include it in your homework. Does mastodon tend to live in places with spruce, or without spruce? Suggest two hypotheses that might explain the observed association.
7. Multi-Time Search
 - a. Hide your previous searches.
 - b. Find all sites with at least 20% *Picea* pollen between 21,000 and 18,000 years ago.
 - c. Find all sites with at least 20% *Picea* pollen between 15,000 and 12,000 years ago.
 - d. Find all sites with at least 20% *Picea* pollen between 10,000 and 7,000 years ago.
 - e. Find all sites with at least 20% *Picea* pollen between 5,000 and 1,000 years ago.
 - f. **QUESTION 7:** Take a screenshot of this map and include it in your homework. Describe the history of *Picea* distributions in eastern North America over the last 21,000 years (i.e. ignore the Rockies and points west). What environmental change(s) might be causing this shift in *Picea* distributions? Given that trees are sessile (immobile) organisms, what processes might allow the observed range shifts?

III. View Data

1. Go find Devil's Lake (WI) again. Click on the site.
2. In the popup window with metadata for Devil's Lake, note that the bottom includes a list of datasets available at the site. (P = pollen; P^ = pollen surface sample; clock = geochronological data; W = water sample; O^ = ostracode surface sample)
3. For the pollen dataset, click on the eyeball at right to view the Devil's Lake pollen dataset. This opens up a new window with more detail about your dataset.
 - a. Samples: A data table. Each row is a different variable and each column is a stratigraphic depth.
 - b. Diagram: Plots a stratigraphic diagram. After clicking on the Settings tab at lower left, Change the Primary Axis to 'PALEON-STEPPS' (This switches the vertical axis from a depth axis to time axis, using an age-depth model developed by the PaleON research project and uploaded to Neotoma. More later about age-depth models.) then click 'Draw.'
 - c. **QUESTION 8:** Take a screenshot of this diagram and include it in your homework assignment.
 - i. What was the most abundant plant taxon at Devil's Lake during the end of the Pleistocene? (i.e. prior to 11,000 years ago)
 - ii. What has been the most abundant taxon during the Holocene? (from 11,000 years ago to present)
 - iii. (Note: The prior two questions implicitly assume that plant with the most abundant pollen in sediments is also the most abundant on the landscape. This isn't necessarily a good assumption; more on pollen representation later.)
 - d. **QUESTION 9:** Using the information in the Chronology tab, how many chronology controls are stored for Devil's Lake? How many of these are radiocarbon dates?
 - i. You might note a systematic offset between the redline (the fitted age-depth model) and the age controls used to fit it. This is a known data-visualization bug

in Neotoma Explorer that is due to the systematic offset between radiocarbon ages and true ages. We'll tackle this in another lab.

- e. **QUESTION 10:** What publications are listed for Devil's Lake? (**Tip:** Go look in Metadata tab)

IV. Download Data

1. In the popup window with metadata for Devil's Lake, note that each dataset has both an eyeball (which you've already clicked on) and a '+' to the right. Click on the '+'. This adds the dataset to a 'Datasets' tray.
2. Find the icon at top for the 'Saved Datasets Tray' and click on it.
3. Hover over the dataset then click on the Save icon to the right. The dataset will be saved as a text file in CSV (comma separated value) format.
4. Open the downloaded CSV file in Excel or a text editor (e.g. Notepad, Wordpad) to look at it.
5. **QUESTION 11:** What is the most abundant taxon at Depth = 1 cm?

V. Explore on Your Own

1. For this last part, explore! This is an unbounded part of the exercise. Using the skills you have learned, search and explore within Neotoma.
2. **QUESTION 12:** What did you find? (~0.5 to 1 page, with a mixture of text and screenshots). There's no wrong answer here; I'm mainly interested in seeing you try out experiments, searches, and visualizations that go beyond the above script. Find something interesting and describe what you find.
3. A few tips:
 - a. **Tip 1:** When searching for taxa, you usually will have better luck if you search at the genus level with all variants than looking for a particular species (see the *Mammut* example above). Neotoma only accepts scientific names for species, so use the internet to find scientific names for the taxa that you're interested in. Neotoma Explorer unfortunately does not yet support recursive taxon searches; i.e. searching at the genus level will not automatically return species-level taxon names. It's on the development to-do list!
 - b. **Tip 2:** Charismatic megafauna are always fun! *Smilodon*, *Mammuthus*, *Megalonyx*,...
 - c. **Tip 3:** some common plants in North American pollen records: *Pinus* (pine), *Betula* (birch), *Quercus* (oak), *Fraxinus* (ash), *Ulmus* (elm), *Tsuga* (hemlock), *Castanea* (chestnut), *Poaceae* (grass), *Ambrosia* (ragweed), ...