CYBERINFRASTRUCTURE RESOURCES IN PALEOBIOLOGY AND PALEOECOLOGY

**NEOTOMA PALEOECOLOGY DATABASE**

**Presenter:** Jack Williams, University of Wisconsin  
http://www.neotomadb.org/

**iDIGBIO**

**Presenter:** Gil Nelson, Florida State University  
https://www.idigbio.org/

**PALEOBIOLGOGY DATABASE**

**Presenter:** Mark D. Uhen, George Mason University  
http://paleobiodb.org

Free Shortcourse at the GSA 2014 Meeting

**DATE:** Tuesday, October 21, 2014  
**TIME:** 2-5pm  
**LOCATION:** Hyatt Regency, English Bay
• Introduction to Neotoma
• Demo – Finding & Accessing Data in Neotoma Explorer
• Demo – Finding & Accessing Data via APIs
• Demo – Finding & Accessing Data via R
• Explore
Q: What is Neotoma?

A: An Open-Access Multiproxy Community Database for the Quaternary-Pliocene
Neotoma Domain

- **Time**: Late Neogene (~last 5 million years)
  - Most records: $10^4$-$10^5$ yrs
- **Space**: North American to Global
- **Taxa**:
  - Pollen,
  - Plant Macrofossils, Mammals,
  - Ostracodes,
  - Testate Amoebae,
  - Insects
  - Etc.

Brewer et al. 2012 TREE
Why NEOTOMA?

Neotoma (packrat, woodrat) collects plants, bones, and other materials, which it deposits in middens, which can be preserved for thousands of years. The middens are cemented and preserved by amberat (dried urine), which contains pollen.

Neotoma collects multiproxy paleodata!

Slide credit: Eric Grimm
Purposes of Neotoma

• To facilitate studies of ecosystem development and response to climate change

• To enable joint analysis of multiproxy datasets to address paleoenvironmental questions that transcend those possible with single-proxy databases.

• To provide the historical context for understanding biodiversity dynamics, including genetic diversity

• To provide the data for climate-model validation

• To facilitate analyses of biostratigraphy

• To provide a safe, long-term, low-cost archive for a wide variety of paleobiological data

• To lower the overall cost of paleodata management
Neotoma Design Concepts

• Partnership between domain scientists and information technology specialists; wherein the science is driving the IT.

• Neotoma offers database infrastructure to specialists in various taxonomic groups, who will not need to develop or even necessarily understand the core information technology, but who can learn to input, update, and extract data through a user-friendly interface and to have control over disciplinary taxonomic issues.

• Neotoma can accommodate virtually any type of fossil data

• Neotoma is a centralized database with virtual constituent databases (e.g. North American Pollen Database, FAUNMAP)

• Constituent database cooperatives may develop individualized websites to frontend the database if they so desire

• Capability for “data stewards” to remotely input and update data

• Access to the data by anyone with an Internet connection
Key Concepts in Neotoma: Sites

Sites are spatial locations from which samples are collected.

<table>
<thead>
<tr>
<th>Sites Table</th>
<th>Kettle Lake</th>
<th>Iron Hills</th>
<th>Parkers Pit</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiteID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SiteName</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LongitudeEast</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LatitudeNorth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LongitudeWest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LatitudeSouth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SiteDescription</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Datasets link together all samples with the same type of measurement from a collection unit at a site.
Key Concepts in Neotoma: Time

*Neotoma stores data about original age controls (e.g. radiocarbon dates) and age models based on those age controls.*

*Alternative age models can be stored & visualized.*

Alternate Age Models

Radiocarbon Dates (Geochron. & ChronControls Tables)

Outlier Radiocarbon Date (Geochronology Table only)
Putting Data in to Neotoma

1) Send data to Data Stewards
   - Pollen: Eric Grimm & Jack Williams
   - Mammals: Jessica Blois & Russ Graham
   - Ostracodes: Alison Smith
   - Diatoms: Don Charles
   - Insects: Alan Ashworth
   - Testate Amoebae: Bob Booth, Paul Hughes

2) Form a Working Group and establish Data Stewards – Webinar Training available

4:30pm today 253-13 Goring et al. Developing Synergies Between Large-Scale Research and Geodatabases: Neotoma and PalEON
Finding and Accessing Data in Neotoma

1. Neotoma Explorer (Map-Based)
   - http://www.neotomadb.org/

2. Neotoma Developer Center (APIs, JSON)
   - http://api.neotomadb.org/doc

3. Neotoma (R statistical package)
   - https://github.com/ropensci/neotoma

4. Tilia
   - http://www.neotomadb.org/pending_data
Neotoma Explorer - Demo

http://www.neotomadb.org/
APIs – Demo

http://api.neotomadb.org/doc/home
neotoma (R) -- Demo

Key Commands:
- get_site
- get_dataset
- get_download

Copy of Demo Script:

Goring et al. in review Open Quaternary
Contact Info

Eric Grimm (pollen, webinar training for stewards, lead PI) grimm@museum.state.il.us
Russ Graham (mammals, lead PI) rgraham@ems.psu.edu
Alan Ashworth (insects) Allan.Ashworth@ndsu.edu
Jessica Blois (mammals) jblois@ucmerced.edu
Bob Booth (testate amoebae) robert.booth@Lehigh.edu
Don Charles (diatoms) dfc63@drexel.edu
Simon Goring (neotoma) goring@wisc.edu
Alison Smith (ostracodes) alisonjs@kent.edu
John (Jack) Williams (pollen, neotoma) jww@geography.wisc.edu
et al.