

# Neotoma Paleoecology Database: Facilitating transparent data curation in the paleosciences

NSF-Geoinformatics  
NSF-EarthCube

Jessica Blois, [jblois@ucmerced.edu](mailto:jblois@ucmerced.edu)

Eric Grimm, [eric.c.grimm@outlook.co](mailto:eric.c.grimm@outlook.co)

Neotoma  
Paleoecology  
Database



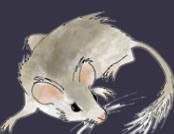
@neotomadb  
@jessicablois

Introductory Virtual Workshop – July 2020  
[www.neotomadb.org](http://www.neotomadb.org)



# Agenda for Today

- 8:30 – 9:30am Neotoma Ecosystem Overview
  - Intro to databases
  - Intro to Neotoma
  - Neotoma data model
  - Neotoma Software Ecosystem
- 9:30 – 10:45am Analyses with Neotoma
  - Overview: Getting data into and out of Neotoma
  - Sandbox time!
- 10:45 – 11:15am BREAK
- 11:15 – 12:15pm Overview of Tilia
- 12:15 – 12:30pm Troubleshoot Tilia installation



# Introduction to paleo-databases: motivations and ecosystem

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



# Paleoecological data: Key characteristics

## Bad News

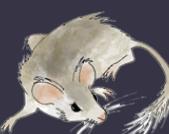
- **'Long Tail'**: Collected in the field & lab by many individuals and scientific teams.
- **Heterogeneity**: Many kinds of measurements & methods
- **Distributed Scientific Expertise** by proxy type, archive type, region, time period, and/or taxonomic group



# Paleoecological data: Key characteristics

## Good News

- **Commonality:** All involve measurements of proxies in various geological *archives* by *depth*, from which we estimate *time*.
- **Long Shelf Life:** specimens & samples collected decades ago are can be re-analyzed



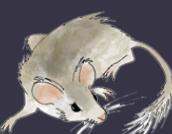
# Community-Curated Data Resources

Social  
Technological

## Socio-Technological Characteristics

- **Shared Mission:** gathering, improving, and sharing data
- Centered on **Communities of Practice**
- **Distributed community governance** to encourage data additions, ensure data quality
- **Centralized IT platform** for collecting, refining, and sharing data
- **Open Data** via multiple outlets
- **Streamlined data uploads** for data and metadata
- **Meso-scale:** bridge between long tail and big data

Neotoma  
Paleoecology  
Database



Paleobiology DB

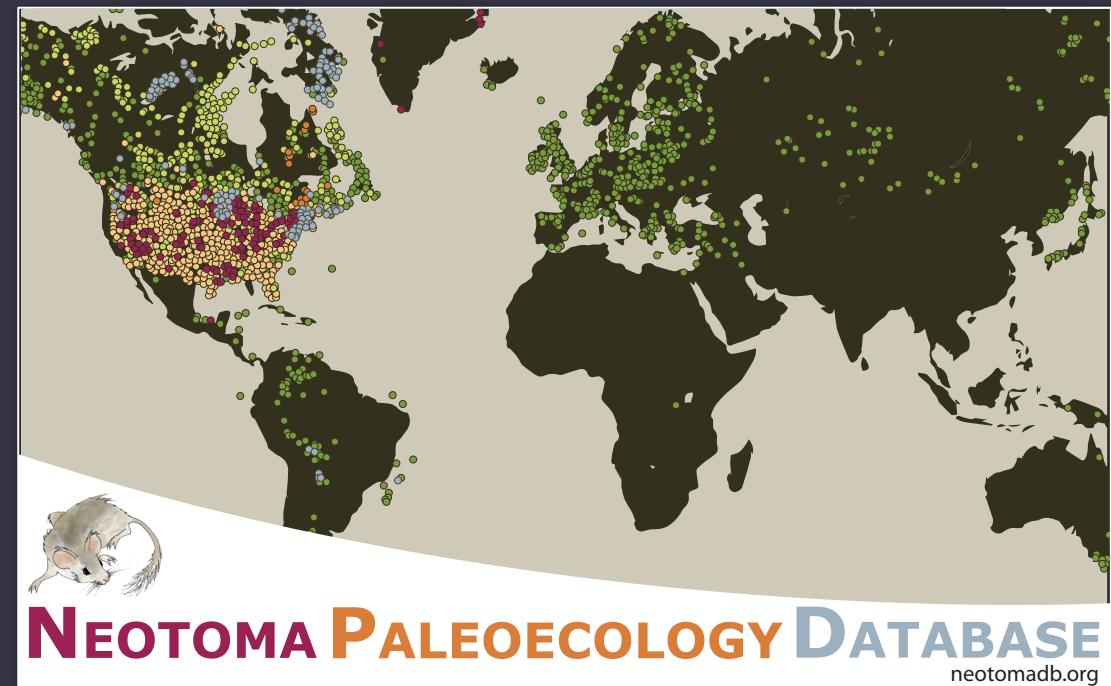


AMERIFLUX

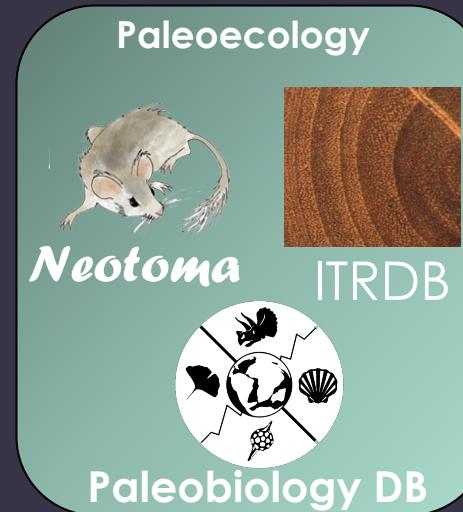
...and many  
others

# Neotoma Paleoecology Database

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



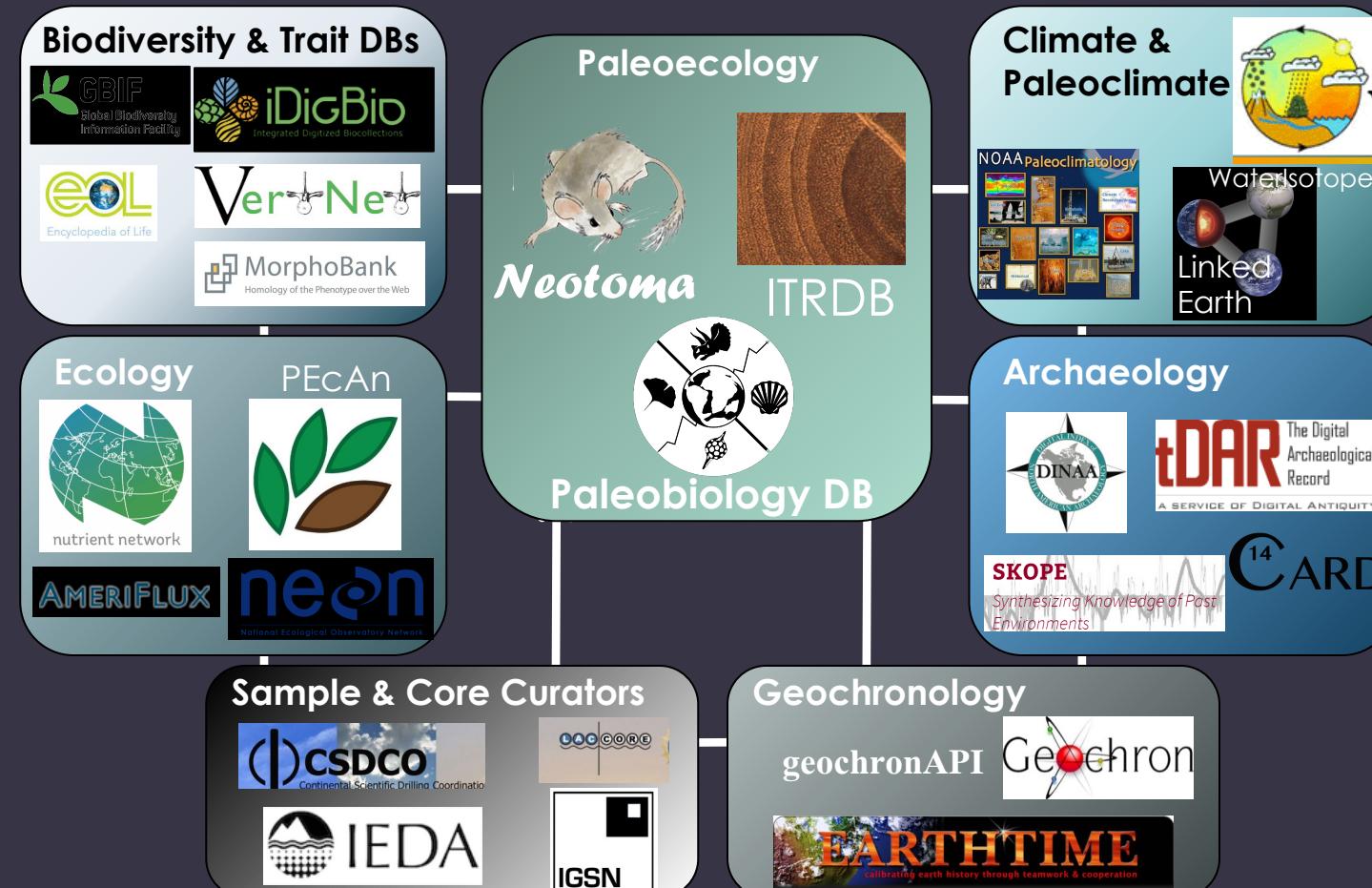
# Growing, Sustaining, and Connecting a Network of Networks



Neotoma  
Paleoecology  
Database



# Growing, Sustaining, and Connecting a Network of Networks

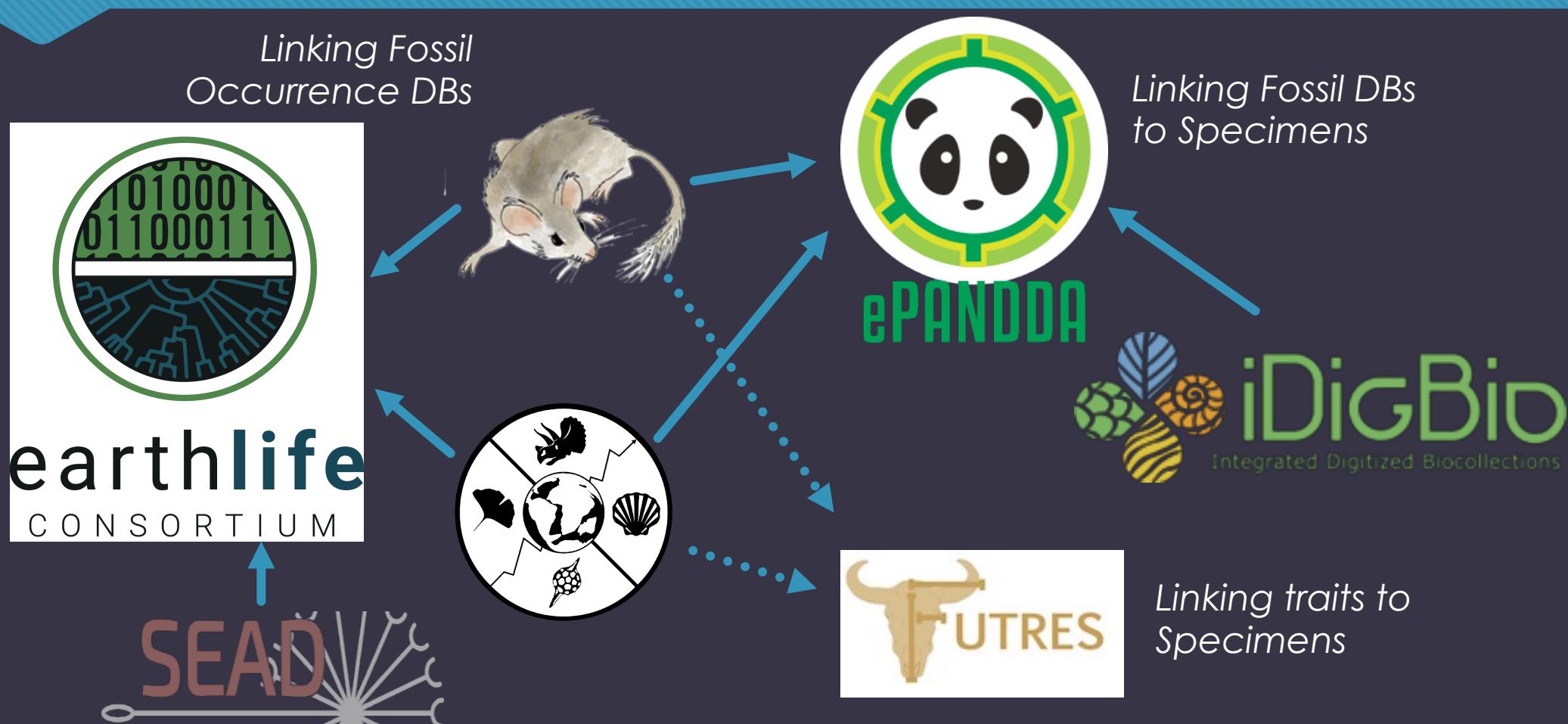


Neotoma  
Paleoecology  
Database

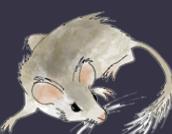


# ELC: Connecting databases through APIs

ELC APIs:  
Documentation & Sandbox  
<http://earthlifeconsortium.org/docs/api-docs.html>

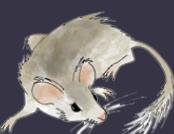
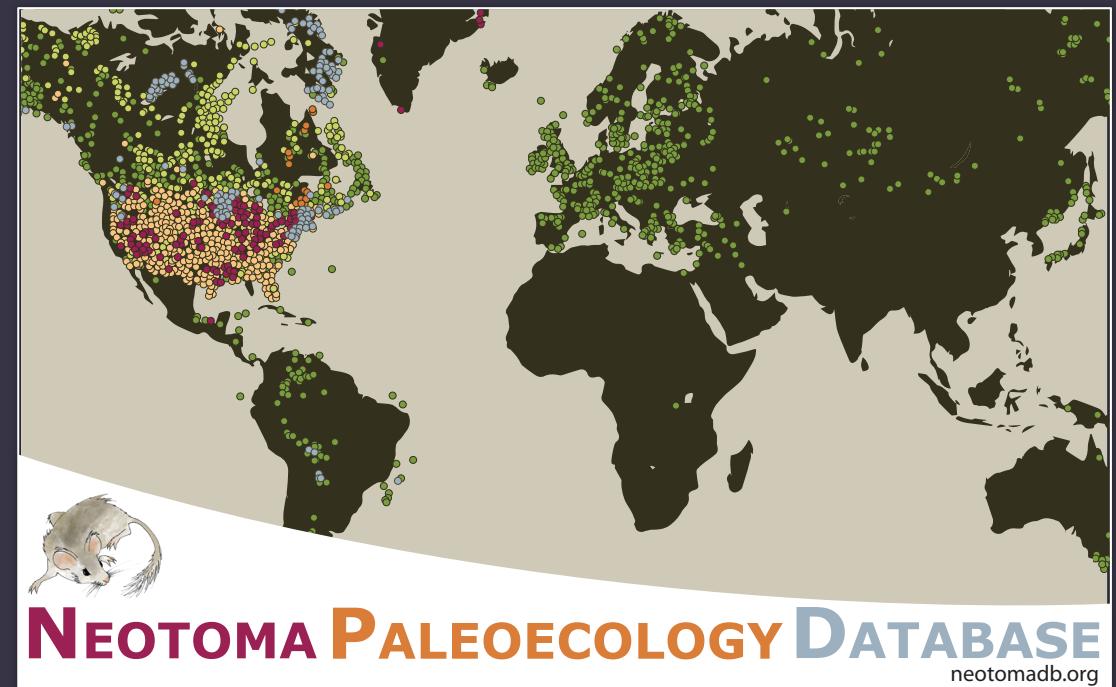


Neotoma  
Paleoecology  
Database



# Neotoma Paleoecology Database

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

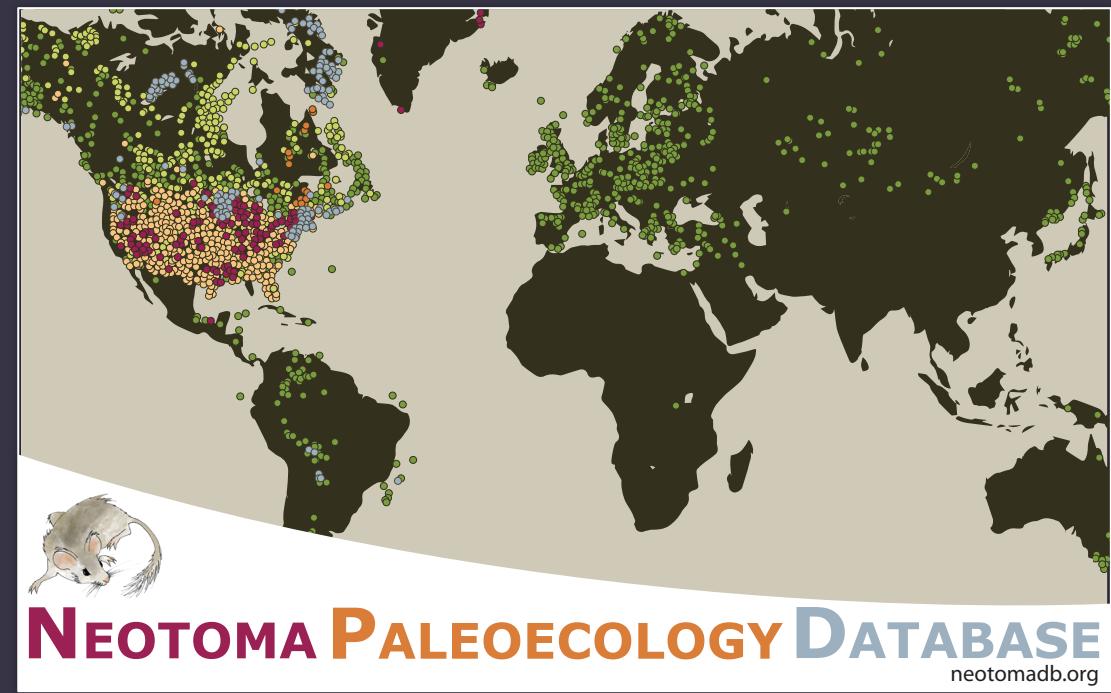


# Neotoma Paleoecology Database

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

## Key Characteristics

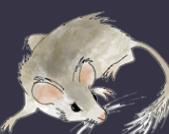
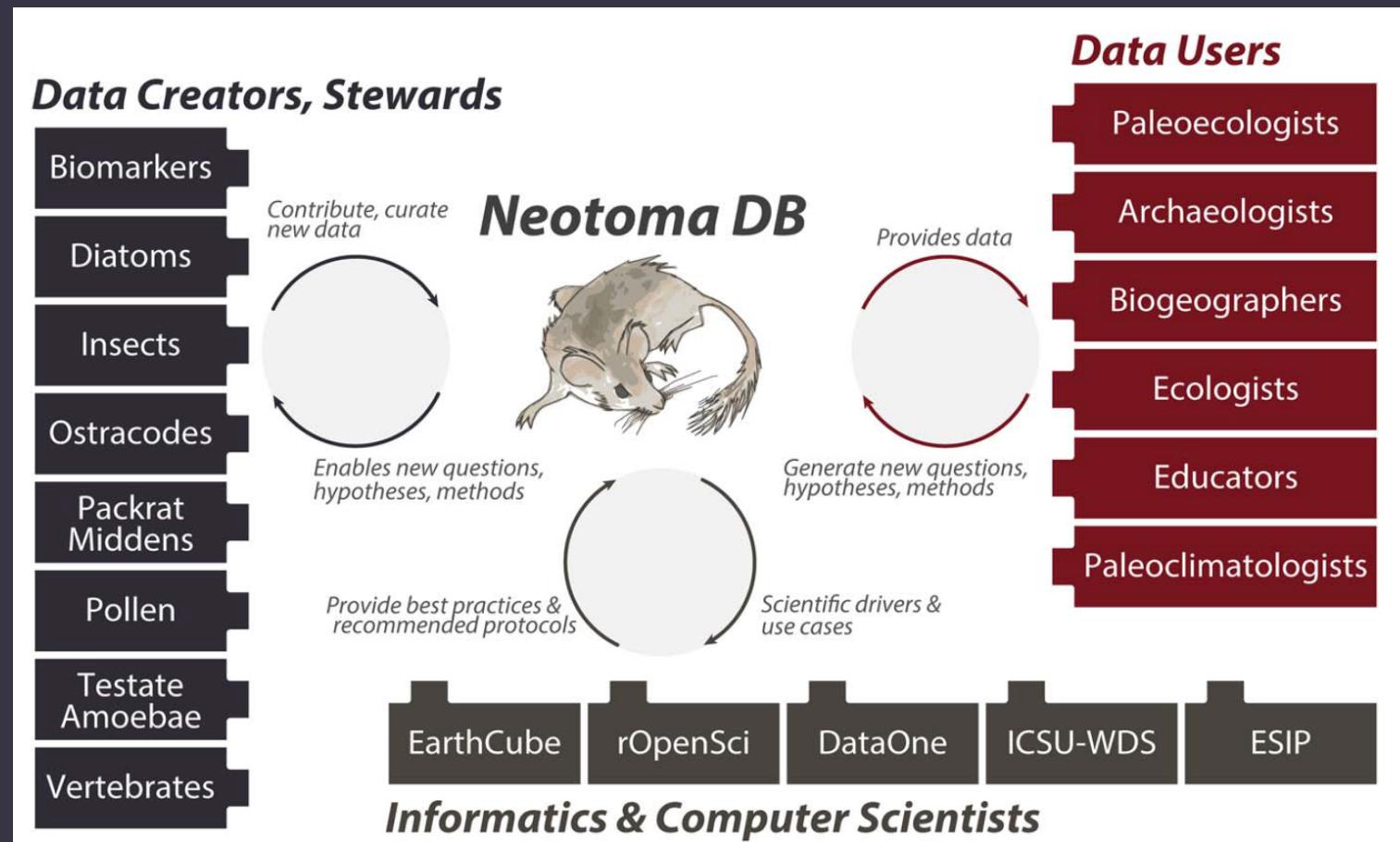
- Open Data
- Curated by Community --> High-Value Data
- Standardized Variable Names & Taxonomy
- Flexible data models for sites, proxies, cores, etc.
- Time: Age Controls and Age Models



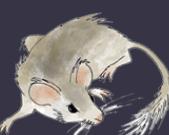
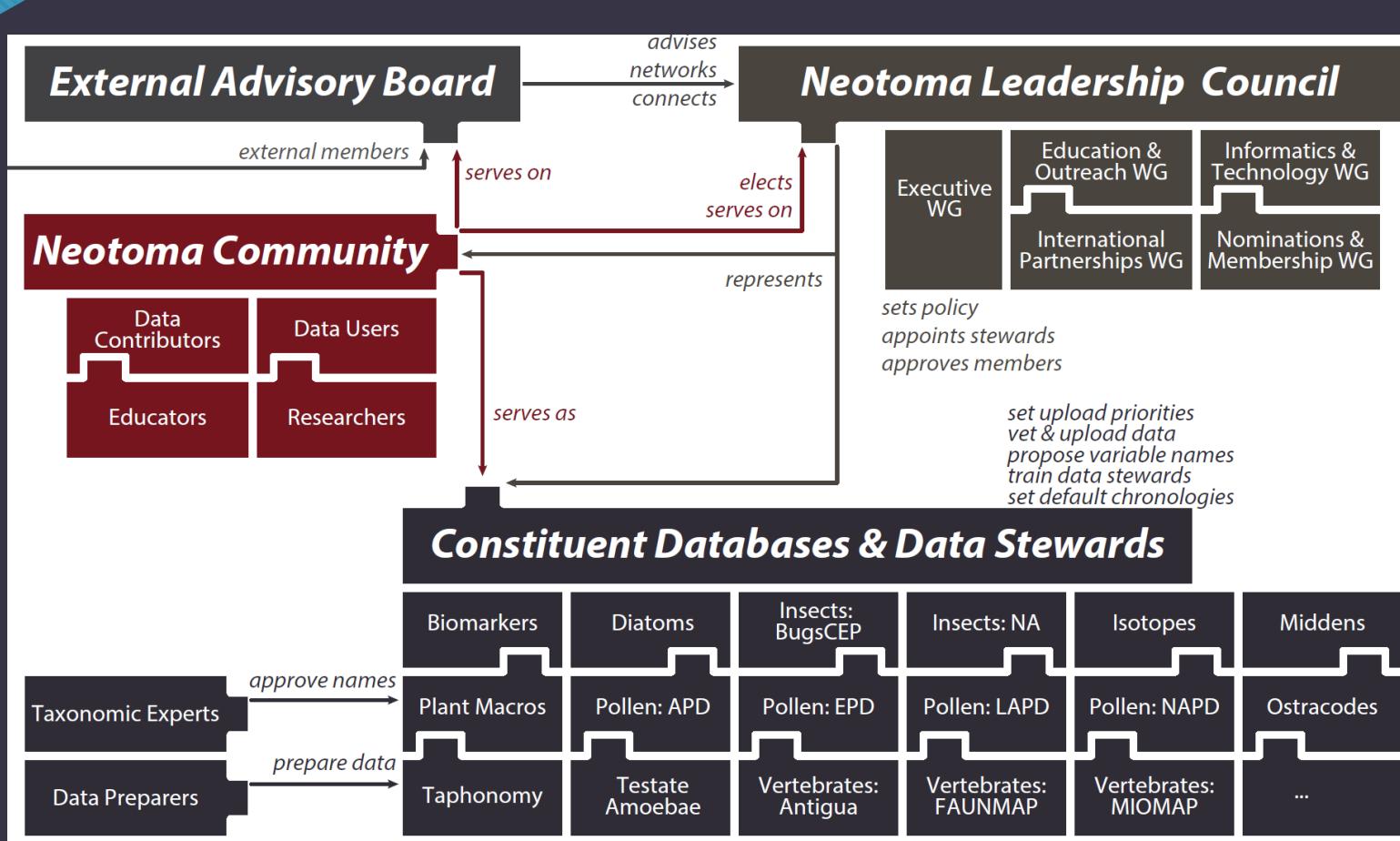
> 7 million individual observations  
from over 38,700 datasets at 18,600 sites



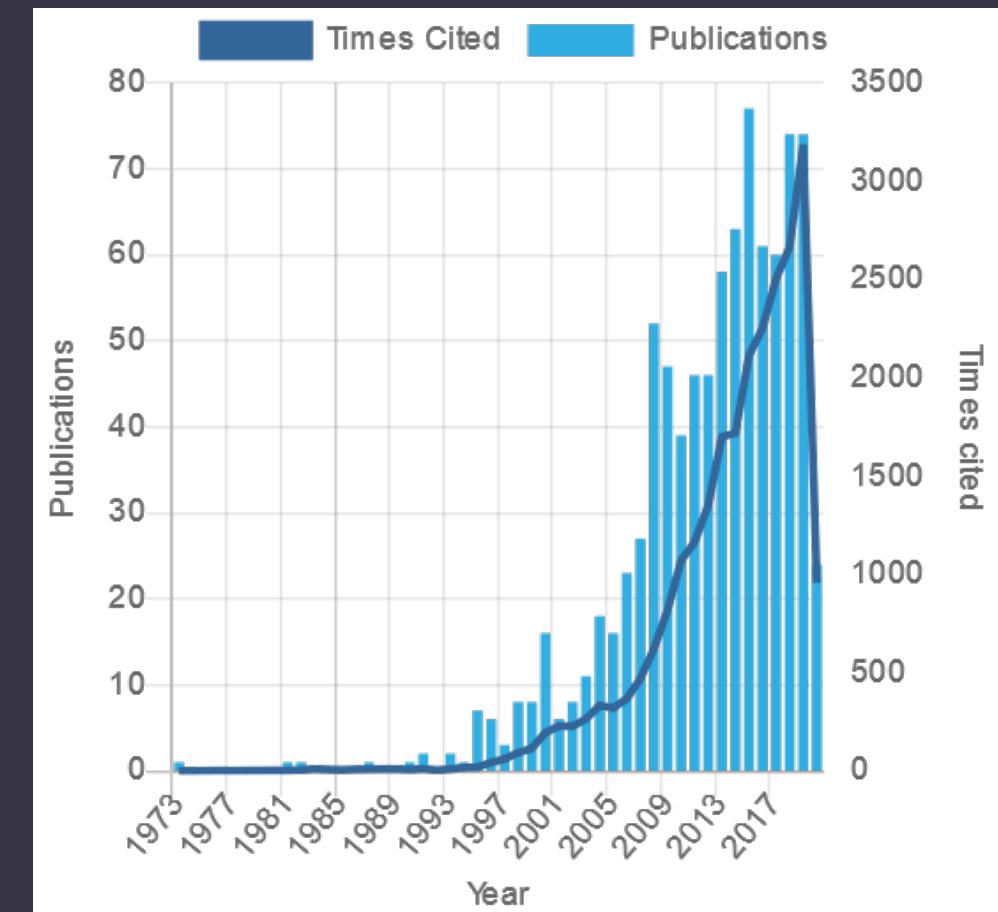
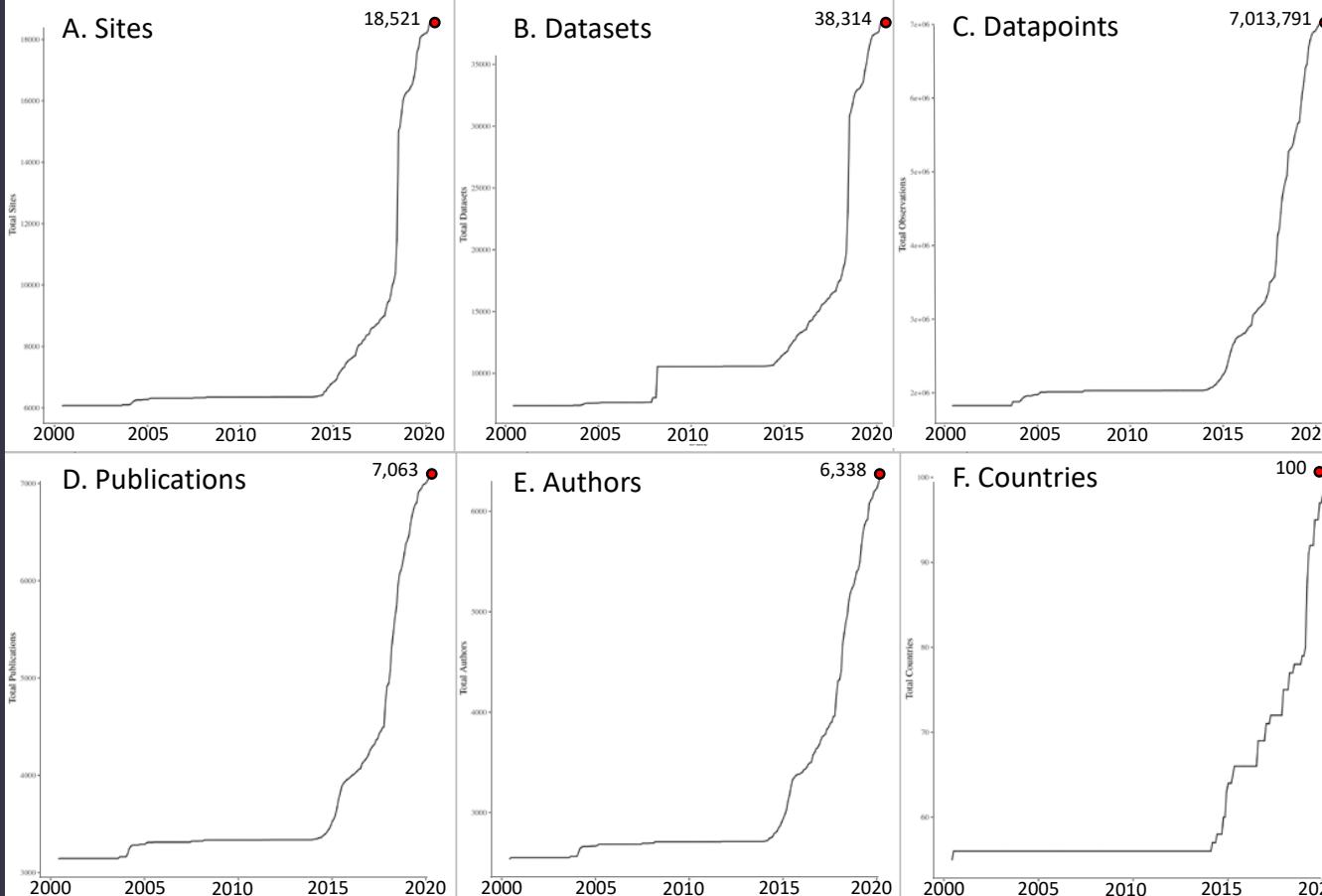
# Community of Neotoma Users



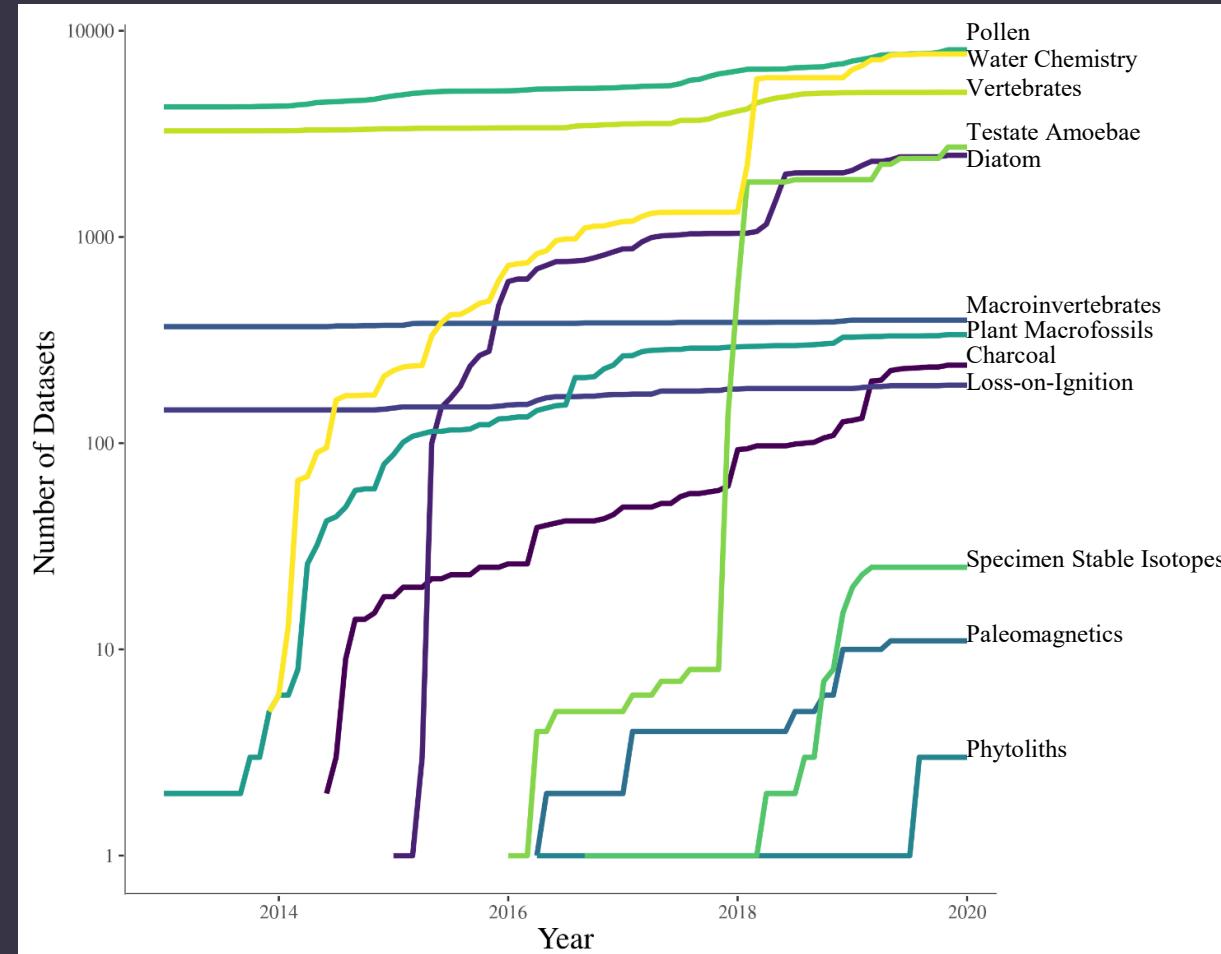
# Distributed & Open Scientific Governance



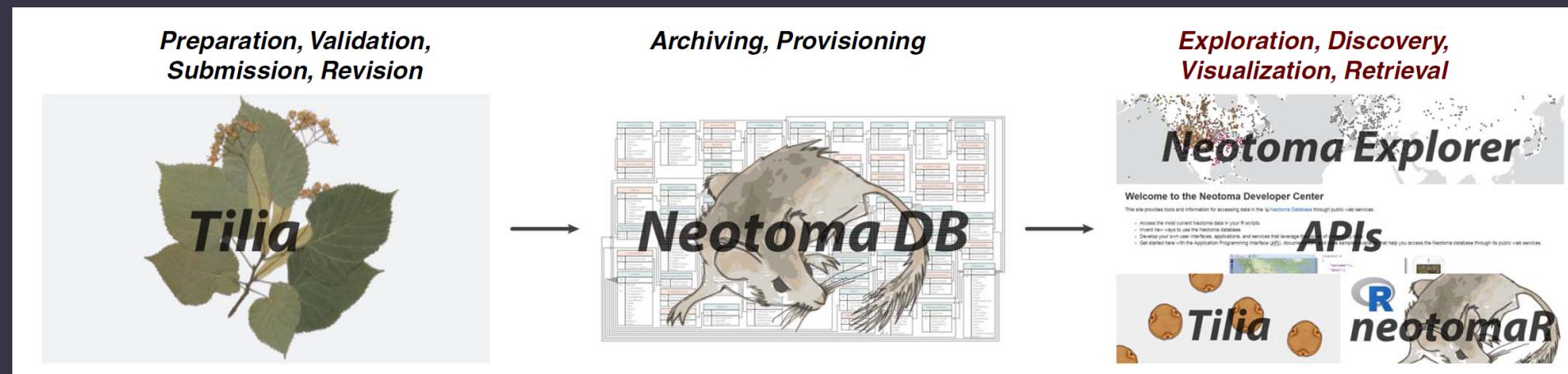
# Database Growth



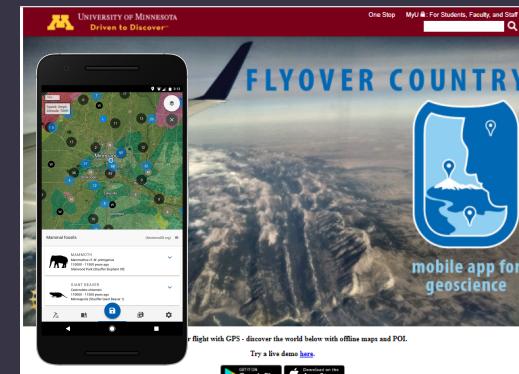
# Data Types & Data Volume



# Neotoma Software Ecosystem: Getting data in and out of Neotoma



## Third-Party Developers



# Getting started with contributing and using Neotoma data

## Uploading data

- Email your raw files to the stewards
- **Submit a Tilia file to the stewards for validation and upload**
- **Become trained as a steward yourself**

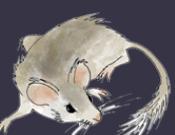
## Accessing data

- Neotoma Explorer:  
<https://apps.neotomadb.org/explorer/>
- Database snapshot:  
<https://www.neotomadb.org/snapshots>
- Tilia
- R (*neotoma*)



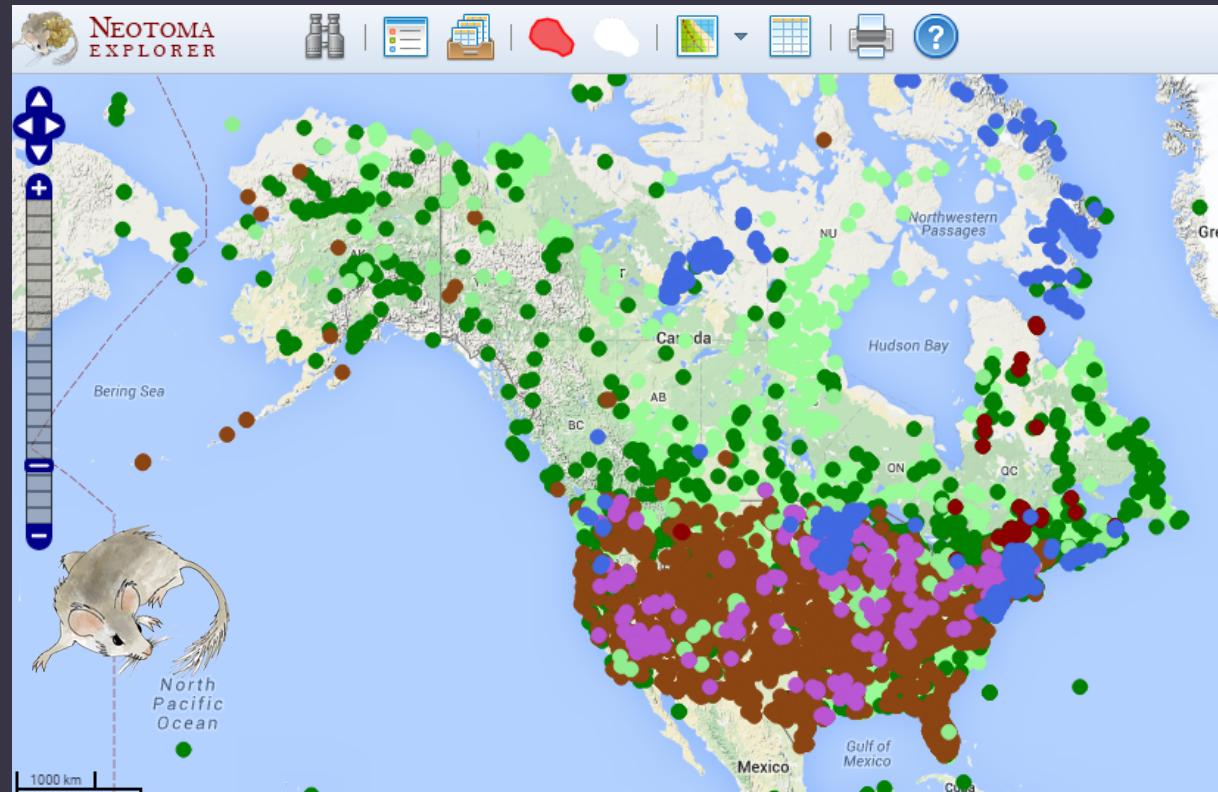
# Tilia (Putting Data into Neotoma)

The screenshot shows the TiliaIT website. At the top center is a detailed illustration of a linden tree branch with green leaves and small yellow flowers. To the left of the main image is a small circular icon with a brown and orange pattern. To the right is a dark rectangular button with a white shopping cart icon. Below the main image is a horizontal menu bar with links: HOME, STORE, DOWNLOAD, CONTACT, FAQ, MAILING LIST, and HELP. In the center of the page is the text "TiliaIT" in a large, bold, green serif font. Below this, there is a thin horizontal green line. At the bottom of the page, there are three navigation links: "Tilia" (in green), "Features" (in green), and "Versions" (in green). The background of the page is white.

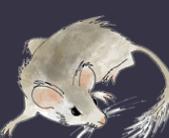


Free Download: <http://www.tiliait.com/download/>  
Help resources: <http://tilia-manual.readthedocs.org/>  
<https://www.tiliait.com/help/>

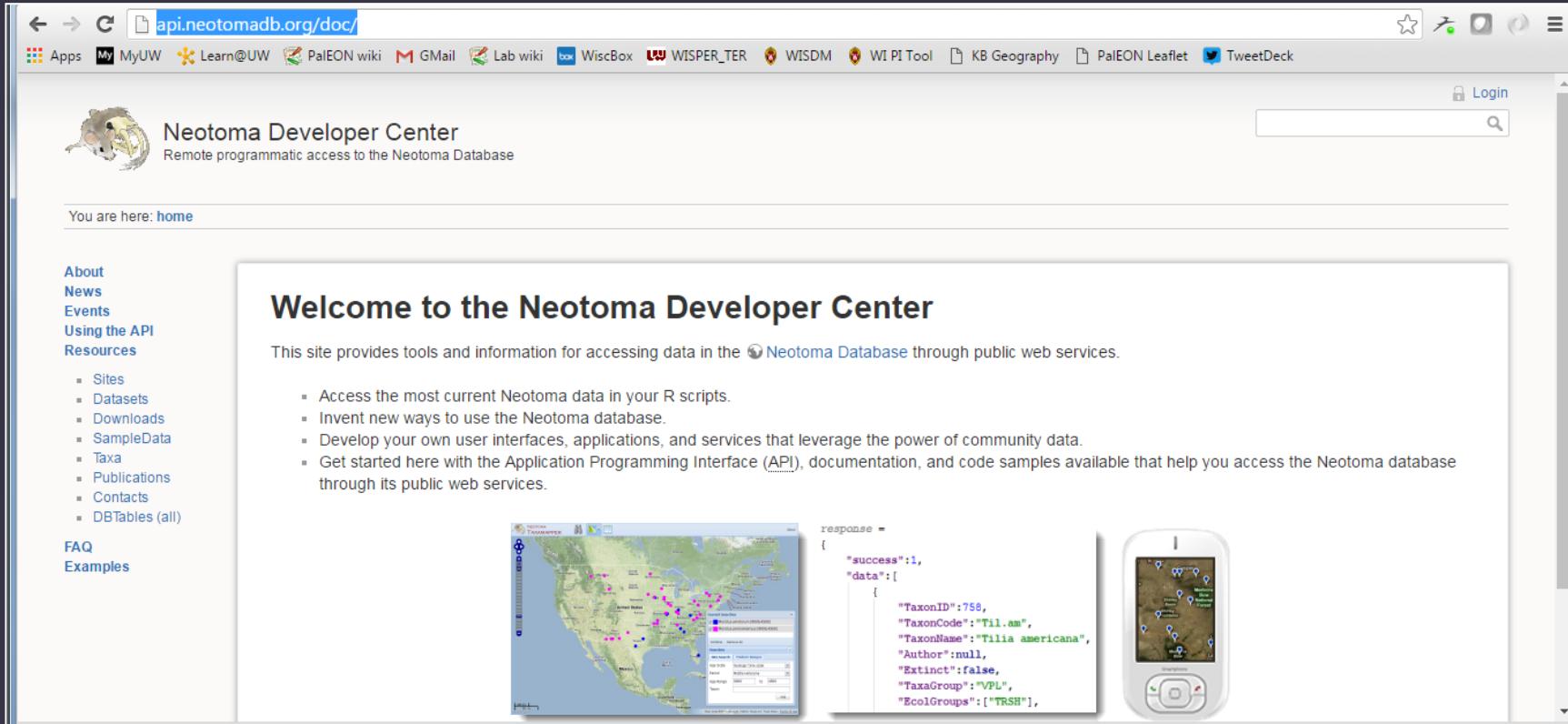
# Neotoma Explorer (Finding/Exploring/Getting Data)



Teaching Exercise, Neotoma Explorer  
<http://serc.carleton.edu/neotoma/activities.html>



# Neotoma APIs (Finding/Exploring/Getting Data)



The screenshot shows a web browser displaying the Neotoma Developer Center at [api.neotomadb.org/doc/](http://api.neotomadb.org/doc/). The page features a header with a mouse icon and the text "Neotoma Developer Center: Remote programmatic access to the Neotoma Database". A sidebar on the left contains links for About, News, Events, Using the API, Resources (with sub-links for Sites, Datasets, Downloads, SampleData, Taxa, Publications, Contacts, and DBTables), FAQ, and Examples. The main content area has a title "Welcome to the Neotoma Developer Center" and a paragraph explaining the site's purpose: "This site provides tools and information for accessing data in the Neotoma Database through public web services." Below this is a bulleted list of four items. To the right of the text are two images: a map of North America with data points and a screenshot of an iPhone displaying a map application.

```
response =  
{  
  "success":1,  
  "data": [  
    {  
      "TaxonID":758,  
      "TaxonCode":"Til.am",  
      "TaxonName":"Tilia americana",  
      "Author":null,  
      "Extinct":false,  
      "TaxaGroup":"VPL",  
      "EcolGroups": ["TRSH"]  
    }  
  ]  
}
```

APIs & Documentation available at:

<http://api.neotomadb.org/doc/>



# Neotoma R

## Porting Neotoma data to R environment

### Key Functions

`get_site`  
`get_dataset`  
`get_download`  
`get_chroncontrol`  
`get_table`  
`get_taxa`

### Other Functions

`get_closest`  
`get_contact`  
`get_publication`

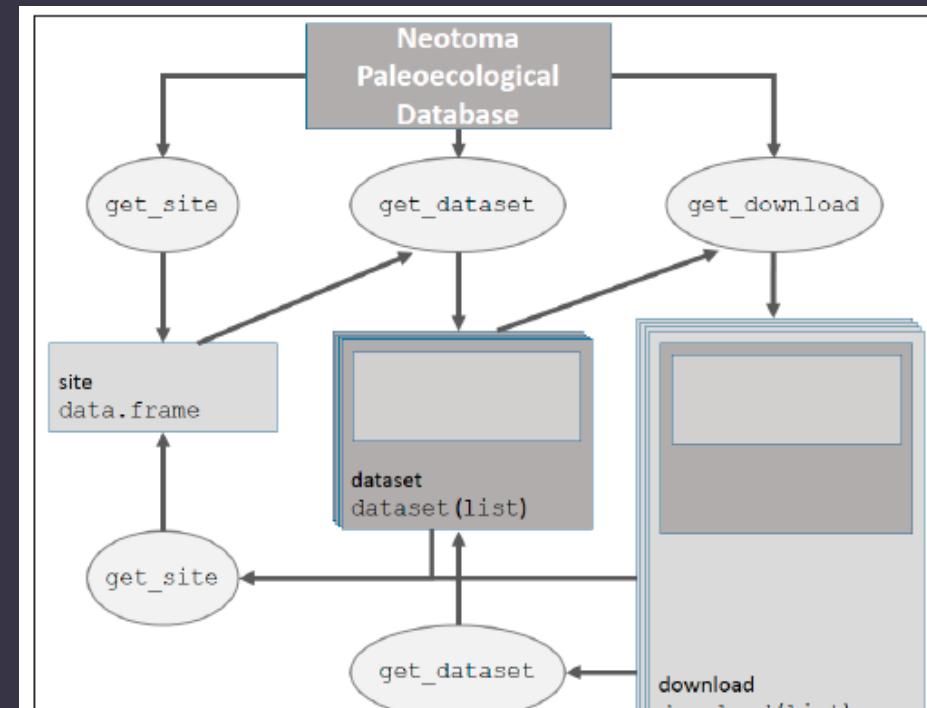


Figure 2: How the main data objects relate to one another in the `neotoma` package, and the helper functions used to move from one data type to another.

GitHub Repository: <https://github.com/NeotomaDB/neotoma>



# Looking Ahead: Goals

## Build Community

- Encourage New Members, Users, Contributors, & Stewards

## Development

- DOIs for Datasets (beta version)
- Embargo Manager
- Link to Other Resources

## Expand Supported Regions & Data Types

- Grow & Establish New Constituent Databases
  - Regions: Australia, Latin America, others?
- Types
  - Organic biomarkers
  - Vertebrate isotopes
  - (sed)ancient DNA



# DOI Landing Pages

Example: <http://data-dev.neotomadb.org/1001>

**NEOTOMA DATASET 1001**  
Hail Lake  
Pollen Dataset

Site Description: Small lake. Surrounding vegetation: *Picea mariana*, *P. glauca*, *Pinus contorta*.

Site Notes: No site notes exists for this site.



Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, IPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012  
Coordinates: | 60.03. -129.02 |

DOI: 10.21233/N31D9W | Neotoma Explorer Link | Download Current Data (JSON) | Download Data As Uploaded (JSON)

**Principal Investigator**  
Cwynar, Les C.  
cwynar@unb.ca

**Chronology**  
Chronology: PALE 1  
Age Type: Radiocarbon years BP  
Prepared by: Mathieu L. Duvall  
Age Range: 60 to 9950

**Publications**  
Cwynar, L.C., and R.W. Spear. 1995. Paleovegetation and paleoclimatic changes in the Yukon at 6 ka BP. Géographie Physique et Quaternaire 49(1):29-35. [DOI: 10.7202/033027ar]

**Other Associated Datasets**  
Datasets at the Same Site  
[8173] Geochronologic  
North American Pollen Database



# Neotoma Resources: Summary

## Neotoma APIs

<http://api.neotomadb.org/doc/>

## ELC APIs

<http://earthlifeconsortium.org/docs/api-docs.html>

## Neotoma Explorer

<http://apps.neotomadb.org/explorer/>

## Flyover Country

<https://flyovercountry.io/>

## Walk-Through Teaching Exercise, Neotoma Explorer

<http://serc.carleton.edu/neotoma/activities.html>

## Tilia

<http://www.tiliait.com/download/>

## neotoma (R package)

CRAN, Goring et al. 2015 Open Quaternary



# Sandbox time!

- There are a variety of resources and tutorials available in the Workshops directory
  - <https://github.com/NeotomaDB/Workshops/>
  - The current workshop is: <https://github.com/NeotomaDB/Workshops/tree/master/CABAH-July2020>

