

Introducción a Neotoma



Neotoma
Paleoecology
Database

Socorro Dominguez dominguezvid@wisc.edu

Simon Goring goring@wisc.edu

bit.ly/LAP

Red Latinoamericana de Polen
9 de mayo, 2025



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON



Socorro Dominguez
Master of Data Science



Simon Goring
Data Scientist II
UW-Madison

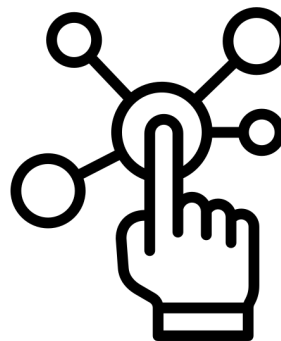
Overview

Introducción a Neotoma

Los Datos en Neotoma

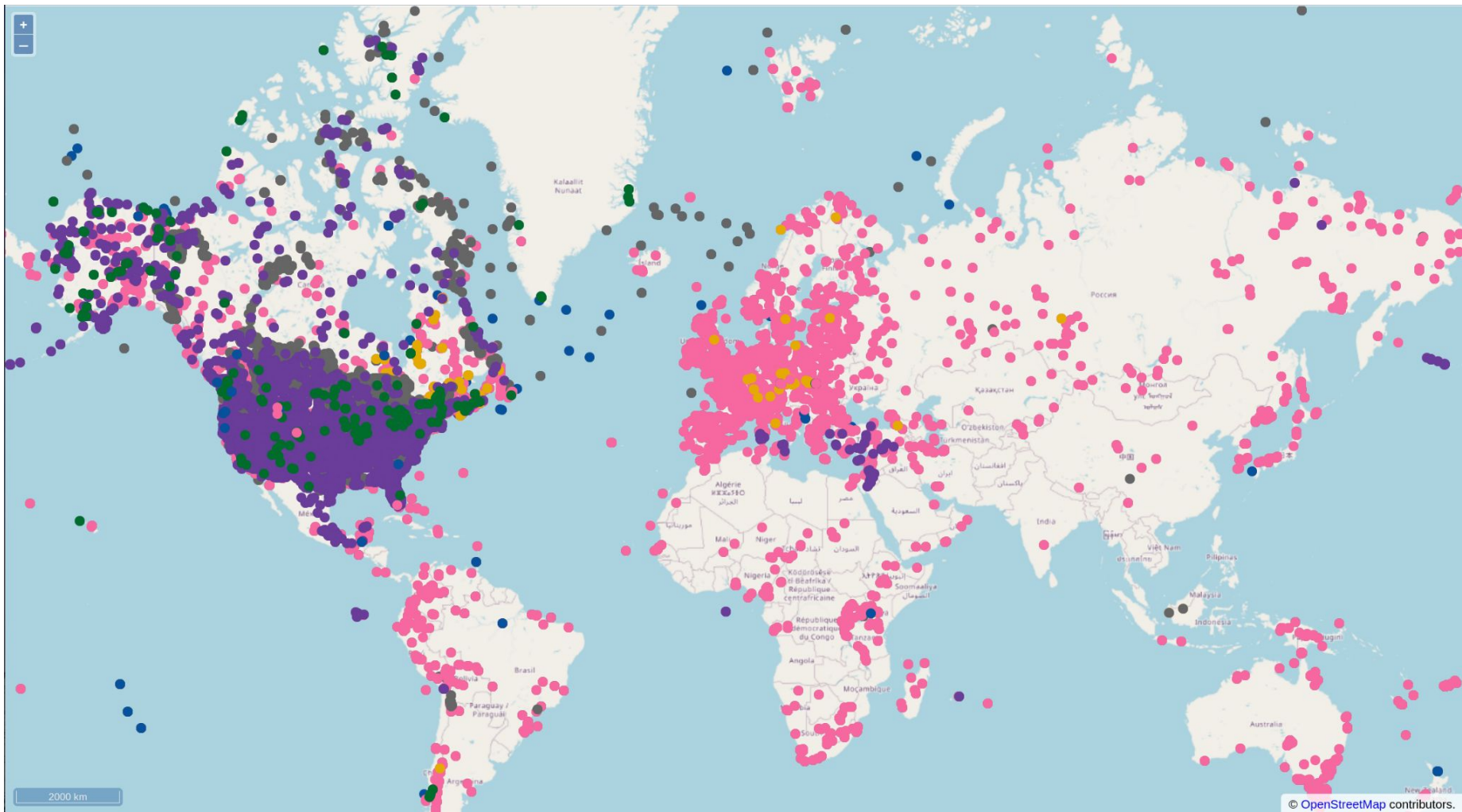
Trabajando directamente con Neotoma

Un “pequeño” ejemplo en R



Created by Eko Purnomo
from the Noun Project

Neotoma significa personas



12,271+

Autores publicados

33,539+

Conjuntos de datos

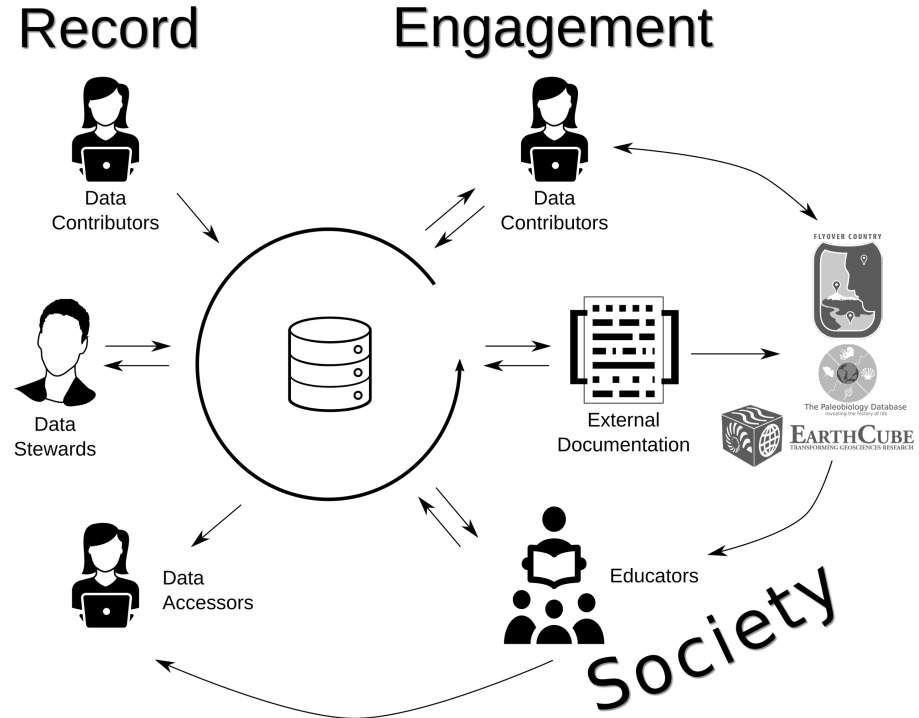
11,156,240+

Observaciones únicas

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., ... Takahara, H. (2018). The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quaternary Research*, 89(1), 156–177.
doi:[10.1017/qua.2017.105](https://doi.org/10.1017/qua.2017.105)



La comunidad de Neotoma



Nuestra página

Aún no la tenemos en español, pero si necesitas ayuda, no dudes en contactarnos.

<https://www.neotomadb.org>

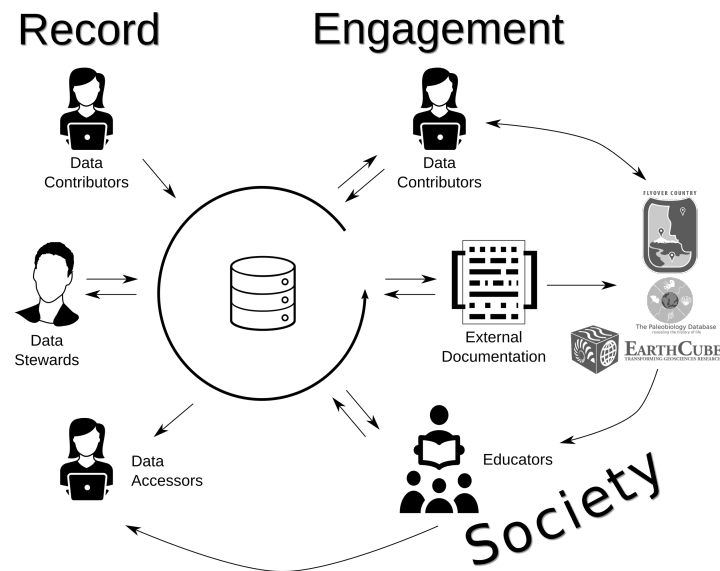
¿De dónde vienen los datos?



Los datos en Neotoma

Tilia (direct entry)

DataBUS [[GitHub](#)]



Bases de datos constituyentes

- Holocene Perspective on Peatland Biogeochemistry
 - FAUNMAP
 - Japanese Pollen Database
 - Neotoma Biomarker Database
 - Sedimentary aDNA Database
 - SISAL
 - **En total, 44 bases de datos que nos constituyen**
-

Extrayendo los datos

Los datos en Neotoma

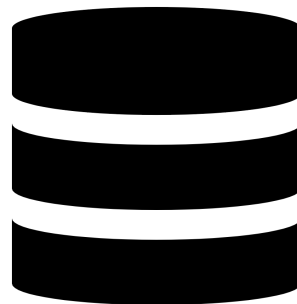


Base de Datos (SQL)

Neotoma API

Explorer

R (neotoma2)



Created by Wilson Joseph
from Noun Project

<https://api.neotomadb.org>

Base de Datos en Postgres

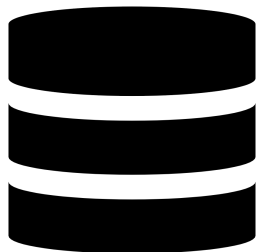
Documentación:

- Manual - <https://open.neotomadb.org/manual>
- Esquema de Datos - <https://open.neotomadb.org/dbschema>

Si quieres aprender o ya sabes SQL, es una gran herramienta

Ideal para casos muy particulares o complejos.

La Base de Datos se respalda cada 4 meses.





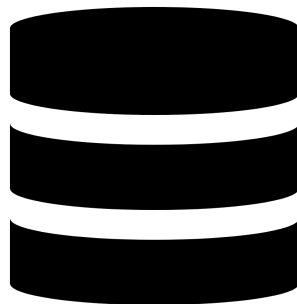
Neotoma Data Pathways

Database (SQL)

Neotoma API

Explorer

R (neotoma2)



Created by Wilson Joseph
from Noun Project

<https://api.neotomadb.org>



Neotoma API

La API genera consultas pre-hechas

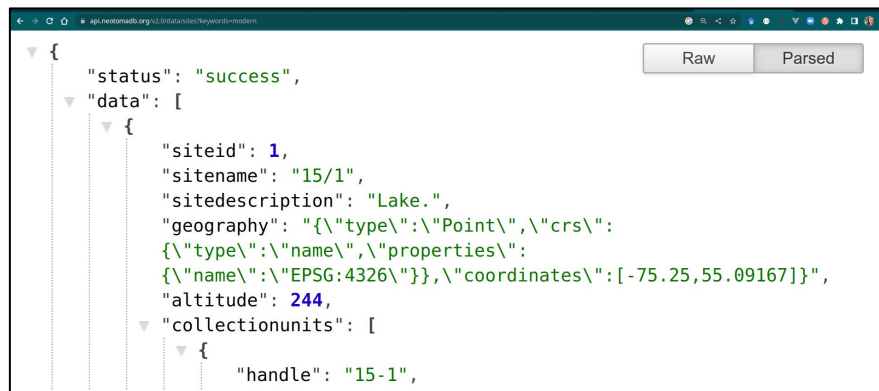
- Obtener sitios por ID
- Obtener un conjunto de datos por nombre del sitio
- Obtener sitios por su ubicación geográfica

<https://api.neotomadb.org/api-docs>

API de Neotoma

Obtiene datos en formato JSON:

`https://api.neotomadb.org/v2.0/data/sites?keywords=modern`



```
{
  "status": "success",
  "data": [
    {
      "siteid": 1,
      "sitename": "15/1",
      "sitedescription": "Lake.",
      "geography": "{\"type\":\"Point\",\"crs\":{\"type\":\"name\",\"properties\":{\"name\":\"EPSG:4326\"}},\"coordinates\":[-75.25,55.09167]}",
      "altitude": 244,
      "collectionunits": [
        {
          "handle": "15-1",
```

Neotoma Data Pathways

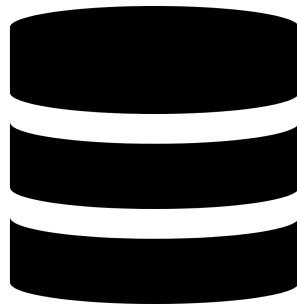


Database (SQL)

Neotoma API

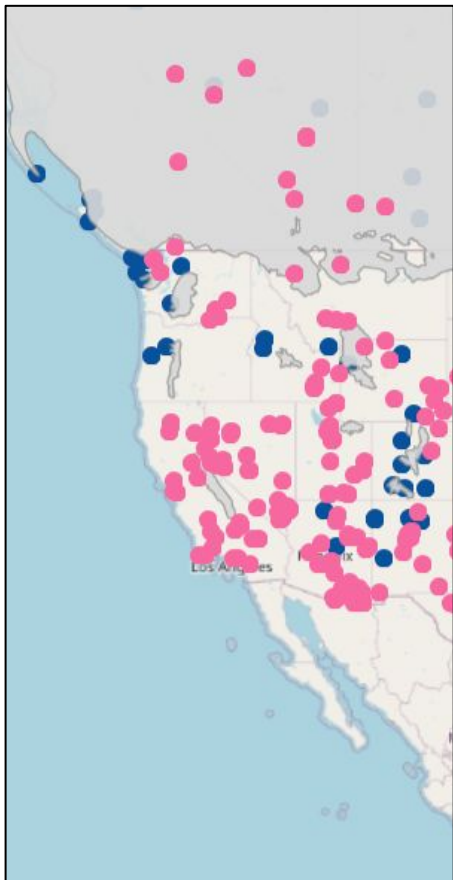
Explorer

R (neotoma2)



Created by Wilson Joseph
from Noun Project

<https://api.neotomadb.org>



Explorer

Una interfaz gráfica

- <https://apps.neotomadb.org/explorer>

Modulos educativos

- <https://serc.carleton.edu/neotoma/index.html>

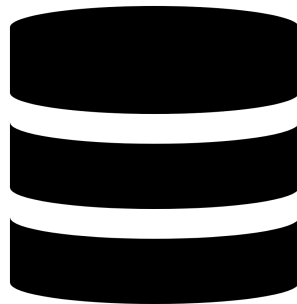
Playlist en YouTube

- <https://www.youtube.com/playlist?list=PLIz1bp5bLPvArVNYUCbyYPmaahEnEysnk>
-

Neotoma Data Pathways



Database (SQL)
Neotoma API
Explorer
R (neotoma2)



Created by Wilson Joseph
from Noun Project



Librería en R: Neotoma2

neotoma2

```
install.packages('neotoma2')
```

- Utiliza métodos geoespaciales
 - Listo para usar con dplyr
 - Métodos ``get_`` & ``set_`` que permiten crear diferentes objetos.
 - Utiliza objetos S4 objects en R - lo que permite una mejor validación de datos.
 - Reestructuración para poder manejar diversos proxies.
-



Fortalezas de `neotoma2`

La estructura de datos en R respeta la estructura de Neotoma:
Sitios -> Colecciones -> Conjuntos de Datos -> Muestras

Los objetos S4 validan los datos

Usamos el estilo de programación de `tidyverse` (e.g., usando las pipas `%>%`) facilitando el uso de otras librerías como: ``stringr``, ``dplyr``, ``lubridate``, etc

¿Cómo trabajamos con neotoma2?

Usamos workflows con LIMITS y OFFSETS:

```
get_sites(gpid="Canada")
```

Dará resultados con 25 observaciones.

Una vez que los pasos analíticos sean concluidos, se puede agregar el parámetro `all_data` evaluado a `TRUE`

```
get_sites(gpid="Canada", all_data = TRUE)
```

¿Cómo trabajamos con neotoma2?

Usamos workflows con LIMITS y OFFSETS:

1.1MB



```
get_sites(gpid="Canada")
```

Dará resultados con 25 observaciones.

Una vez que los pasos analíticos sean concluidos, se puede agregar el parámetro `all_data` evaluado a `TRUE`

3.9MB!



```
get_sites(gpid="Canada", all_data = TRUE)
```

Devuelve más de 100 registros

Ejemplos de Trabajo “Cortos”

Simple Vignette:

[https://open.neotomadb.org/Current Workshop/simple workflows.html](https://open.neotomadb.org/Current%20Workshop/simple%20workflows.html)

Complex Vignette:

[https://open.neotomadb.org/Current Workshop/complex workflows.html](https://open.neotomadb.org/Current%20Workshop/complex%20workflows.html)