

# What can APIs do for you?

Wendy Gross<sup>1</sup>, C. Morrill<sup>1</sup>, E. Gille<sup>1,2</sup> and E. Shepherd<sup>1,3</sup>

As the spatial and temporal resolution of scientific datasets and the culture of data sharing grow, more data on past global changes are now available than ever before. Efficiently discovering, downloading, and integrating data into analyses is critical for making full use of this surge of information. Researchers have long used Graphical User Interfaces<sup>1</sup>, or GUIs, to manually search and download data, but Application Programming Interfaces, or APIs, can handle these tasks in a more automated fashion. In fact, APIs are the behind-the-scenes conduit for many different types of information we use every day, from weather forecasts within smartphone apps to flight schedules aggregated on travel websites.

APIs are the technological backbone that let two computer programs communicate over the internet. Each API defines a set of rules that specify the parameters by which requests (or "calls") can be made, as well as the format of the computer-readable information that is provided in response. World Data System repositories such as the World Data Service for Paleoclimatology (WDS-Paleo)<sup>2</sup>, PANGAEA<sup>3</sup>, and Neotoma<sup>4</sup> provide APIs, as do other data and information sources such as publishers (e.g. Springer, Wiley), publication databases (e.g. CrossRef, Web of Science), domain-specific databases (e.g. Global Biodiversity Information Facility), and analysis tools (e.g. ArcGIS).

A request to an API is usually written in the form of a specially-formatted web address, or Uniform Resource Locator (URL). Scientists can call an API by simply entering such a URL into their web browser or by incorporating calls to an API in data analysis code. These new capabilities open more automated ways of first finding and accessing information, and then integrating information, both from different sources, and with different analysis tools.

The ability to discover and download information programmatically via an API, as opposed to manually through a GUI, increases efficiency and diminishes the possibility for human error. For example, APIs make it easier to repeat a search (perhaps to find newly archived datasets or updates to existing datasets) or to gather information quickly for initial data exploration (perhaps to identify certain geographical areas or parameter types with sufficient amounts of data for an analysis). APIs can also make it more efficient to search multiple data providers. While requests must be structured to match the requirements of a specific API, they can be a useful way to locate data across several repositories. In fact, the federated data search provided by the WDS-Paleo API uses the Neotoma API to retrieve information about datasets.

API requests can be integrated with many programming languages (e.g. Python, R, MatLab), effectively creating a pipeline of data to tools for analysis. For example, functionality to access data from the PANGAEA, and Neotoma data repositories via their APIs exists in some Python<sup>5</sup> and R<sup>6,7</sup> packages (e.g. Goring et al. 2015) and the WDS-Paleo also provides example API requests in these languages<sup>8</sup>. Incorporating API requests into a scientific workflow promotes reproducibility and repeatability of research. Encoding all steps of the workflow from data discovery and download to processing and analysis provides complete documentation of the research methods, including the criteria used to select datasets. Some APIs also perform analysis directly: for example in the geospatial ecosystem, ArcGIS APIs<sup>9</sup> and Open Geospatial Consortium APIs<sup>10</sup> perform geospatial analysis, including image analysis and feature classification.

While APIs provide the ability to access formatted information directly from computer systems of authoritative sources, thereby

streamlining data access and analysis, there are still some obstacles to integrating data from different repositories or sources seamlessly (e.g. EarthLife Consortium API<sup>11</sup>; Uhen et al. 2021). For example, interoperability and reusability of paleoenvironmental data also require enhanced common standards and workflows for metadata and data reporting (Bothe et al. 2021; Khider et al. 2019; Morrill et al. 2021; Williams et al. 2018). These improvements, in concert with technological advances such as APIs, will accelerate discovery from the many decades of data collection by the paleo community.

## AFFILIATIONS

<sup>1</sup>National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Boulder, CO, and Asheville, NC, USA

<sup>2</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, CO, USA

<sup>3</sup>Riverside Technology, Inc, Fort Collins, CO, USA

## CONTACT

Wendy Gross: [paleo@noaa.gov](mailto:paleo@noaa.gov)

## REFERENCES

Bothe O et al. (2021) *PAGES Mag* 29: 59

Goring S et al. (2015) *Open Quat 1: Art 2*

Khider D et al. (2019) *Paleoceanogr Paleoclimatol* 34: 1570-1596

Morrill C et al. (2021) *Paleoceanogr Paleoclimatol* 36: e2020PA004193

Uhen MD et al. (2021) *Front Biogeogr* 13: e50711

Williams J et al. (2018) *PAGES Mag* 26: 50-51

## LINKS

<sup>1</sup>For example: [ncei.noaa.gov/access/paleo-search/pangaea.de](https://ncei.noaa.gov/access/paleo-search/pangaea.de), [apps.neotomadb.org/explorer](https://apps.neotomadb.org/explorer)

<sup>2</sup>[ncei.noaa.gov/access/paleo-search/api](https://ncei.noaa.gov/access/paleo-search/api)

<sup>3</sup>[ws.pangaea.de](https://ws.pangaea.de)

<sup>4</sup>[api.neotomadb.org/api-docs](https://api.neotomadb.org/api-docs)

<sup>5</sup>[github.com/pangaea-data-publisher/pangaeapy](https://github.com/pangaea-data-publisher/pangaeapy)

<sup>6</sup>[docs.ropensci.org/pangaeaar](https://docs.ropensci.org/pangaeaar)

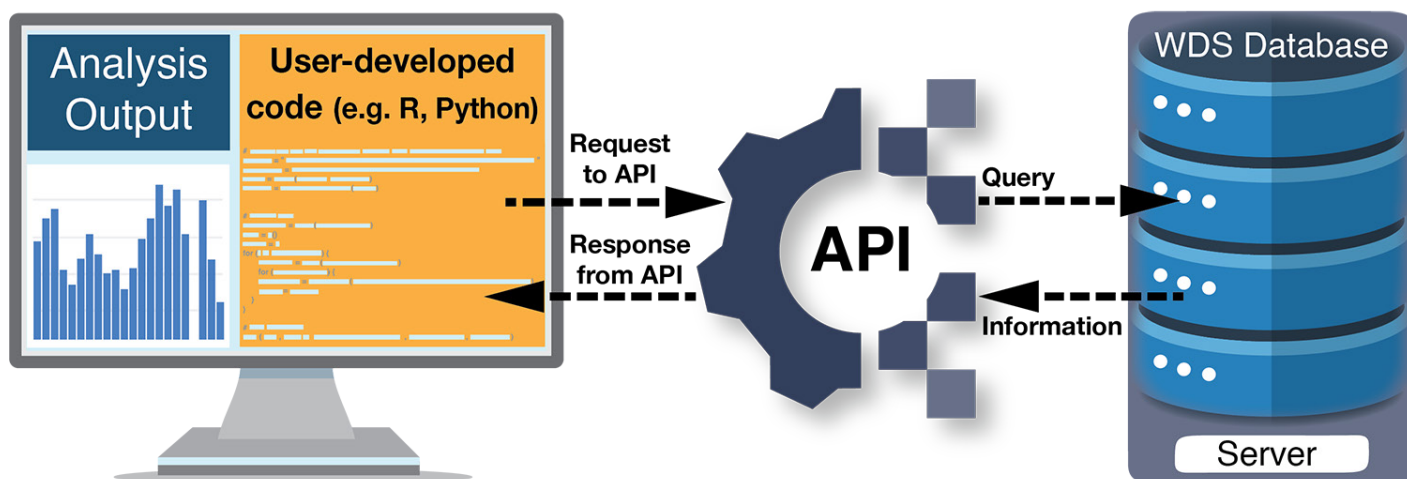
<sup>7</sup>[cran.r-project.org/web/packages/neotoma](https://cran.r-project.org/web/packages/neotoma)

<sup>8</sup>[ncei.noaa.gov/access/paleo-search/api#examples](https://ncei.noaa.gov/access/paleo-search/api#examples)

<sup>9</sup>[developers.arcgis.com/rest](https://developers.arcgis.com/rest)

<sup>10</sup>[ogc.org/resource/demos-archive](https://ogc.org/resource/demos-archive)

<sup>11</sup>[earthlifeconsortium.org](https://earthlifeconsortium.org)



**Figure 1:** Example workflow for programmatically using a World Data Service (WDS; e.g. WDS-Paleo, PANGAEA, Neotoma) API. Starting with user-developed code (yellow box), information from a data provider is discovered and accessed via the API, and then incorporated into a scientific analysis. See the WDS-Paleo API Tutorial<sup>8</sup> for sample code in R and Python.