Drive-Thru Data: Using NLM APIs to Access Information Fast

Mike Davidson, MLS

National Library of Medicine

National Institutes of Health

U.S. Department of Health & Human Services

U.S. National Library of Medicine

After this session, you should be able to…

- Explain…
  - what an API is,
  - how APIs can help you interact with systems, and
  - why users might choose to use APIs.
- Describe the basic mechanics of using an API
- Identify…
  - some of NLM's APIs, and
  - know when a specific NLM API would be useful.

The World of NLM

Literature
- PubMed
- PMC
- Bookshelf

Consumer Health
- MedlinePlus

Terminology
- MeSH
- RxNorm
- UMLS

Molecular Biology
- Nucleotide
- Protein
- SRA

Drugs and Chemicals
- DailyMed
- PubChem

Other
- ClinicalTrials.gov

And many more…

Poll: Which categories of NLM products do you use?

- Literature
- Consumer Health
- Terminology
- Molecular Biology
- Drugs and Chemicals
- Other

Example: Health info for patients in EHRs

Example: Finding clinical trials for cancer patients

Example: Author nationality trends in PubMed

Example: Author nationality trends in PubMed (continued 1)

Example: Author nationality trends in PubMed (continued 2)

NIH U.S. National Library of Medicine

What do these projects have in common?

- Each uses information from NLM resources…
  - …but **not** those resources' websites!
- They need a different type of access:
  - Outside of a web browser
  - With limited (or zero) direct human interaction
  - To information in a specific format.
  - To information as data.

The solution?
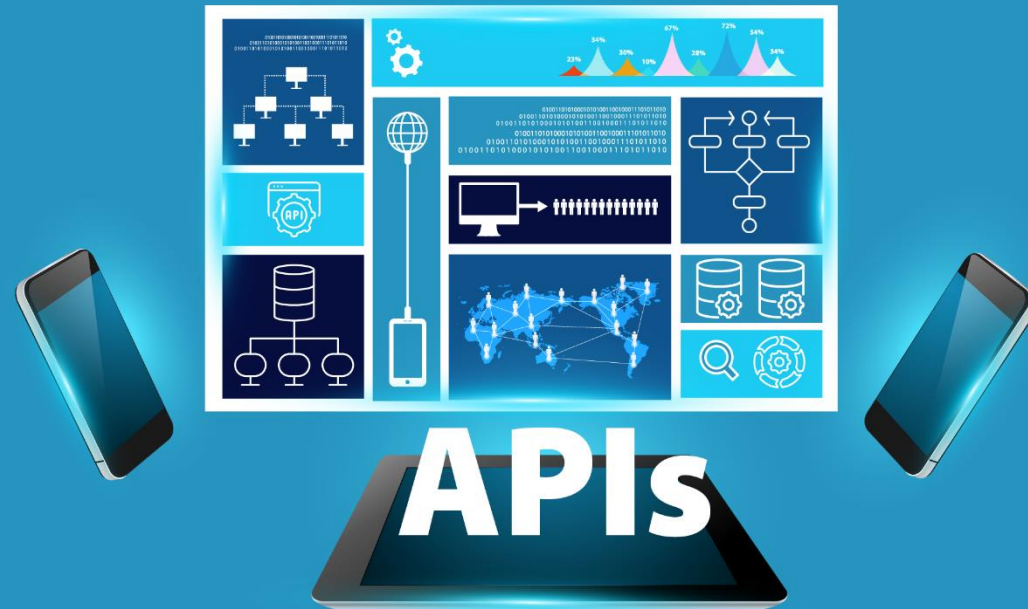
Application Programming Interfaces (APIs)

Poll: What is your experience with APIs?

- Use them all the time!
- Use them periodically.
- Have used them in the past.
- Know about them, but haven't used them.
- This is all new to me!

What is an API?

- A set of protocols for contacting a remote system and making requests.

- Designed to be used "programmatically," not directly by humans.

- APIs typically include:
  - a server, and
  - a set of rules for making requests (or "calls") to that server

A Drive Thru for Data

Why are APIs useful?

- API calls can be built-in to programs/applications.
  - Data can be requested/retrieved much faster
  - Less need for human intervention
- Some APIs offer more options for data retrieval.
  - Retrieval in specialized formats
  - Retrieval of otherwise unavailable data.

How (many) APIs work

- The way you access the API is via a URL
- The specific URL you use includes the address of the API you're using, plus the details of your request
- What information you get back depends on how you construct the URL.

The two parts of (many) API requests

The base URL

Indicates which API you're using

Some parameters

The details of what you're asking for

The Base URL

- The address of the API server
- Specific to each individual API
- Some examples:
  - MedlinePlus: https://wsearch.nlm.nih.gov/ws/query
  - E-utilities: https://eutils.ncbi.nlm.nih.gov/entrez/eutils/
  - MeSH RDF: https://id.nlm.nih.gov/mesh

Parameters

- Parameter options are specific to the API in question
  - Actual parameters are specific to each request
- Can include things like:
  - Search strings
  - Results restrictions
  - Formatting options
  - etc.
- An example:
  - db=pubmed&id=1602668&retmode=xml&rettype=full

U.S. National Library of Medicine

Building an API URL: MedlinePlus

- Start with the Base URL for MedlinePlus API
  - https://wsearch.nlm.nih.gov/ws/query

Building an API URL: Part Two

- Determine your parameters
  - Language: English or Spanish?
    - db=healthTopics
  - Search query: What are you looking for?
    - term=acid+reflux
  - Other options: How many results?
    - retmax=5

Putting it all together: MedlinePlus

Base URL

Parameters

Database

Search Query

# of Results

| https://wsearch.nlm.nih.gov/ws/query | db=healthTopics | term=acid+reflux | retmax=5 |

https://wsearch.nlm.nih.gov/ws/query**?**db=healthTopics&term=acid+reflux&retmax=5

What we get
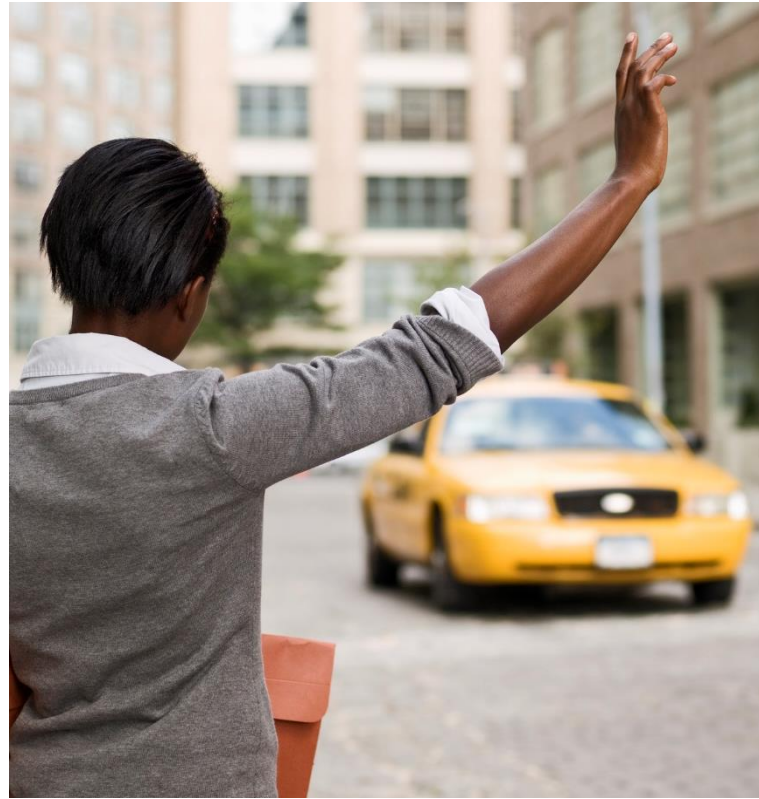
You Need a Car

What kind of car?

What if I don't know how to drive?

Learn to drive…

…or find a driver!

Poll: Any programming experience?

- R
- Python
- C++
- Shell scripting (Linux/Unix)
- PHP/JavaScript
- MatLab
- Other (tell us in chat)
- None yet!

What's on the menu? NLM data!

Choosing the right API

- Remember! Different APIs for different purposes!
- When deciding to use an API, first question: does it have what I need?
- If a resource has multiple APIs, may serve different data in different formats.

MedlinePlus

- MedlinePlus Web Service
  - Retrieves MedlinePlus Health Topics in XML
  - Can help embed MedlinePlus content on a webpage
- MedlinePlus Connect
  - Integrated into Electronic Health Records
  - Used primarily by EHR vendors/developers

PubMed

- E-Utilities
  - Access 35+ NCBI databases, including PubMed
  - Best way to access PubMed via API
- Literature Citation Exporter
  - Converts PMIDs/PMCIDs into citation strings
- Citation Matcher
  - Programmatic access to PubMed Citation Matcher

PMC/Bookshelf

- E-utilities (again)
  - Access metadata and (some) full-text
  - Uses same syntax as E-utilities for PubMed
- OAI-PMH/OAI-PMH LitArch
  - Full-text from PMC/Bookshelf Open Access subsets
  - Uses industry standard for online digital repositories

Medical Subject Headings (MeSH)

E-utilities (yet again)

MeSH RDF

RxNorm

ClinicalTrials.gov

Poll: Which API is most interesting?

- MedlinePlus Web Service/MedlinePlus Connect
- E-utilities (PubMed, MeSH, PMC)
- Other PubMed APIs
- OAI-PMH (PMC, Bookshelf)

- MeSH RDF
- RxNormAPI
- ClinicalTrials.gov
- Something else
- Still not sure why I would use an API…

To recap: when should I use APIs?

- Working in a programming environment
- Need NLM data in a machine-readable format
- Need up-to-date data quickly/on-demand
- Have specific things to search for/request

When should I *not* use APIs?

- When you're NOT programming!
- When APIs aren't an option
- When exploring/browsing a resource
- When you need **all** of the data

Bulk Downloads

https://www.nlm.nih.gov/databases/download/data_distrib_main.html

Where to go next?

- Learn about programming or find a programmer
  - Online courses
  - Library Carpentry
  - Ask around!
- Think about your project
  - What do you know?
  - What do you need to know?
- Find the right API for you

NLM Data Discovery

Read the documentation!

- Tells you what the API can and can't do
- Instructions on syntax, formatting requests
- Provide guidelines for usage
- May include example API calls

U.S. National Library of Medicine

Poll: What else do you need?

- More examples of APIs in action
- More info on available NLM APIs
- Help with programming
- Something else? – In the chat!

Questions?

U.S. National Library of Medicine

How to Get Your MLA CE

Close survey

Thank you for completing this evaluation survey. Your feedback will support efforts to improve future Network of the National Library of Medicine (NNLM) training sessions.

If your training session   offered Continuing Education credit from the Medical Library Association and you would like to claim it, please follow these instructions. See Step 4 for enrollment code.

1. Go to www.medlib-ed.org

2. Login (for information on how to create a free MEDLIB-ED account, *visit MLA's FAQ page for further help*).

3. Click My Learning on the blue bar near the top of the page

4. Enter the following enrollment code in the appropriate field:

   ███████ (please copy)

5. Click Redeem, then Claim

If you have questions or run into problems with MEDLIB-ED, please email MEDLIB-ED@mail.mlahq.org

Copy the code from the evaluation survey