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
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

![Clinical Trial Search](image)

96 results have been found. The results are based on: Search parameters

<table>
<thead>
<tr>
<th>Status</th>
<th>Matching Criteria</th>
<th>Study Title</th>
<th>Conditions</th>
<th>Interventions</th>
<th>Eligibility Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruiting</td>
<td>Age is matching</td>
<td>AN OPEN-LABEL, MULTICENTER, RANDOMIZED PHASE 3 STUDY OF FIRSTLINE ENCOREAFENIB PLUS CETUXIMAB WITH OR WITHOUT CHEMOTHERAPY VERSUS STANDARD OF CARE THERAPY WITH A SAFETY LEAD-IN OF ENCOREAFENIB AND CETUXIMAB PLUS CHEMOTHERAPY IN PARTICIPANTS WITH METASTATIC BRAF V600E/MUTANT COLORECTAL CANCER</td>
<td>Neoplasms</td>
<td>Encoreafenib</td>
<td>show more</td>
</tr>
<tr>
<td>Recruiting</td>
<td>Age is matching</td>
<td>An Open-label Phase 1 Study to Evaluate Drug-Drug Interactions of Agents Co-Administered With Encoreafenib and Bevacizumab in Patients With BRAF V600-mutant Unresectable or Metastatic Melanoma or Other Advanced Solid Tumors</td>
<td>Advanced Solid Tumors</td>
<td>Metastatic Melanoma</td>
<td>show more</td>
</tr>
<tr>
<td>Recruiting</td>
<td>Age is matching</td>
<td>Detection and Metabolic Characterization in Dopa PET/CT of Brain Metastases MRI</td>
<td>Brain Metastases/ MRI</td>
<td>F-Dopa PET/CT</td>
<td>show more</td>
</tr>
</tbody>
</table>

Locations:
- Phoenix | Mayo Clinic - Phoenix Oncology Pharmacy/AZ
- Phoenix | Mayo Clinic Hospital - AZ
- Scottsdale | Mayo Clinic in Arizona - Scottsdale/ AZ
- Beverly Hills | Tower Hematology Oncology Medical Group (THO) - CA
- Los Angeles | Ask Hospital of USC - CA

- Orange | UC Irvine Health - CA
- Aurora | University of Colorado Hospital - Aurora Cancer Pavilion (ACP) - CO
- Chicago | University of Illinois at Chicago - IL
- Saint Paul | Regions Cancer Care Center - MN
- Saint Paul | HealthPartners Specialty Center-Eye Care - MN
- Angers | CHU Angers - Angers | Institut de Cancérologie de l'Ouest
Example: Author nationality trends in PubMed
Example: Author nationality trends in PubMed (continued 1)
Example: Author nationality trends in PubMed (continued 2)
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:
  - 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
Building an API URL: MedlinePlus

- Start with the Base URL for MedlinePlus API
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

<table>
<thead>
<tr>
<th>Base URL</th>
<th>Database</th>
<th>Search Query</th>
<th># of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://wsearch.nlm.nih.gov/ws/query">https://wsearch.nlm.nih.gov/ws/query</a></td>
<td>db=healthTopics</td>
<td>term=acid+reflux</td>
<td>retmax=5</td>
</tr>
</tbody>
</table>

What we get

This XML file does not appear to have any associated metadata or annotations. The document tree is shown below.

```xml
<nlUnsearchResult>
  <term>acid reflux</term>
  <file>vic_lhn.info</file>
  <server>pubmed</server>
  <count>12</count>
  <restart>0</restart>
  <retmax>5</retmax>
</nlUnsearchResult>

This document is not available online, but the following text is included:

Acid	Reflux

Acid reflux happens when a muscle at the end of your esophagus does not close properly. This allows stomach contents to leak back, or class="qt1" class="reflux" class="span">
</doc>

Reflux

You may feel a burning in the chest or throat called heartburn. Sometimes, you can taste stomach fluid in the back of the mouth. If you have these symptoms more than twice a week, you may have GERD. You can also have GERD without having heartburn. Your symptoms could include a dry cough, asthma symptoms, or trouble swallowing.

Anyone, including infants and children, can have GERD. If not treated, it can lead to more serious health problems. In some cases, you might need medicines or surgery. However, many people can improve their symptoms by:

- Avoiding alcohol and spicy, fatty or acidic foods that trigger heartburn.
- Eating smaller meals.
- Not eating close to bedtime.
- Losing weight if needed.
- Wearing loose-fitting clothes.
- Taking steps to keep your weight in check.

A National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) publication about heartburn.

You may also want to speak with your doctor or pharmacist about:

- Medicines that can help with GERD.
- Support groups for people with GERD.

If you have GERD, your symptoms could include:

- Heartburn, a burning sensation in the center of your chest.
- Acid reflux, the transmission of stomach acid to your throat.
- Indigestion, a burning sensation in your upper abdomen.

If you have questions about your symptoms or treatment options, speak with your doctor.

For more information, visit the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) website.
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)

This XML file does not appear to have any style information associated with it. The document tree is shown below.

MeSH RDF
RxNorm
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

National Library of Medicine Data Distribution

The NLM Data Distribution program is the preferred access point for bulk downloading of the data for the products listed below. Downloading and use of these datasets is free of charge and implies agreement to the Terms and Conditions. For questions or assistance regarding the data distributed within this program, either visit the NLM Support Center or email at custserv@nlm.nih.gov.

MEDLINE/PubMed

NLM produces a baseline set of MEDLINE/PubMed citation records in XML format for download on an annual basis. Each day, NLM produces update files that include new, revised, and deleted citations.

MeSH

Medical Subject Headings (MeSH) is a hierarchically-organized terminology for indexing and cataloging of biomedical information. It is used for the indexing of PubMed and other NLM databases.

Catalog Record Data

Bibliographic records for books, journals and other materials from NLM's collections. Includes downloads of CatFile, CatFilePlus, and SerefFile data in NLMXML, MARCXML and MARC 21 formats.

Searching for other resources?

Data Discovery

NLM's Data Discovery is a catalog of NLM products and services with information and links to product information, data downloads, APIs, and other programmatic interfaces.

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

Data Discovery

Access, explore, and build with datasets and APIs from the National Library of Medicine.

- Full Catalog: Browse all datasets, APIs, and other assets in the Data Discovery catalog.
- List of NLM Products: A comprehensive listing of products and services at the National Library of Medicine.
- Annual MeSH Processing: Overview and reports detailing annual changes made to MeSH.

About Data Discovery at the NLM  Terms and Conditions
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
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?
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 was 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:

   [Redacted] (please copy)

5. Click Redeem, then Claim

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

Copy the code from the evaluation survey