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

Mike Davidson, MLS
Office of Engagement and Training
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

<table>
<thead>
<tr>
<th>Status</th>
<th>Matching Criteria</th>
<th>Study Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruiting</td>
<td>Age is matching</td>
<td>AN OPEN-LABEL, MULTICENTER, RANDOMIZED PHASE 3 STUDY OF FIRSTLINE ENCRAFENIB PLUS CETUXIMAB WITH OR WITHOUT CHEMOTHERAPY VERSUS STANDARD OF CARE THERAPY WITH A SAFETY LEAD-IN OF ENCRAFENIB AND CETUXIMAB PLUS CHEMOTHERAPY IN PARTICIPANTS WITH METASTATIC BRAF V600MUTANT COLORECTAL CANCER</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 Encrafenib and Bevacizumab in Patients With BRAF V600-Mutant Unresectable or Metastatic Melanoma or Other Advanced Solid Tumors</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>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Interventions</th>
<th>Eligibility Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoplasms</td>
<td>Encrafenib</td>
<td>show</td>
</tr>
<tr>
<td>Chelatplatin</td>
<td>Inflecan</td>
<td>show</td>
</tr>
<tr>
<td>Leuzovinin</td>
<td></td>
<td>show</td>
</tr>
<tr>
<td>Advanced Solid Tumors</td>
<td>Metastatic Melanoma</td>
<td>show</td>
</tr>
<tr>
<td>Isocetadexamthrine</td>
<td>Cefamandole</td>
<td>show</td>
</tr>
<tr>
<td>Metopranolone</td>
<td>Moxizolide</td>
<td>show</td>
</tr>
</tbody>
</table>

Locations:
- Phoenix | Mayo Clinic - Phoenix Oncology Pharmacy | Arizona
- Phoenix | Mayo Clinic Hospital - Arizona
- Scottsdale | Mayo Clinic in Arizona - Scottsdale | Arizona
- Beverly Hills | Tower Hematology Oncology Medical Group (THO) | California
- Los Angeles | Keck Hospital of USC | California
- Orange | UC Irvine Health | California
- Aurora | University of Colorado Hospital - Anschutz Cancer Pavilion (ACCP) | Colorado
- Chicago | University of Illinois at Chicago | Illinois
- Saint Paul | Regions Cancer Care Center | Minnesota
- Saint Paul | HealthPartners Specialty Center-Eye Care | Minnesota
Example: Author nationality trends in PubMed
Example: Author nationality trends in PubMed (cont.-1)
Example: Author nationality trends in PubMed (cont.-2)
What do these projects have in common?

- Each uses NLM information…
  - …but **not** NLM 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

**Base URL**


**Parameters**

<table>
<thead>
<tr>
<th>Database</th>
<th>Search Query</th>
<th># of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>db=healthTopics</td>
<td>term=acid+reflux</td>
<td>retmax=5</td>
</tr>
</tbody>
</table>

**Complete URL**

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:

```
<summary>
  <item Name="MeshYearIntroduced" Type="String">1969</item>
  <item Name="MeshSoupCode" Type="String">1969</item>
  <item Name="MeshSupplement" Type="String">1969</item>
  <item Name="MeshSuperterm" Type="String">1969</item>
  <item Name="MeshDescription" Type="String">1969</item>
  <item Name="MeshDefinition" Type="String">1969</item>
  <item Name="MeshSupplement" Type="String">1969</item>
  <item Name="MeshSuperterm" Type="String">1969</item>
  <item Name="MeshDescription" Type="String">1969</item>
  <item Name="MeshDefinition" Type="String">1969</item>
</summary>
```

MeSH RDF

MeSH RDF API

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP</td>
<td>The SPARQL 1.1 endpoint returns RDF results and graphs. See the <a href="#">open page</a>.</td>
</tr>
</tbody>
</table>

Endpoints:

- **lookup**
  - Search for Descriptors: Qualifier pairs, also known as Subject Headings.
  - Search for Descriptive Terms.

- **lookup/descriptor**
  - Returns all allowed Qualifiers for a Descriptor.

- **lookup/detail**
  - Lookup some details for a descriptor.

- **lookup/year**
  - Returns the current status of versions years in 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

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
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.

NLM Products and Services
A comprehensive listing of products and services at the National Library of Medicine.

Featured: PubMed Citations
A baseline set of MEDLINE/PubMed citations records in XML format for download on an annual and daily basis.

About Data Discovery at the NLM
Data Discovery is a platform providing access to datasets from selected NLM resources, as well as information and links to other NLM resources. Users can explore, filter, visualize, and export data in a variety of formats, including JSON.

Terms and Conditions
Most datasets on Data Discovery follow the NLM Terms and Conditions for use. Individual datasets may have specific licenses or additional terms of use. Contact the NLM Help Desk for questions about specific datasets.
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?