Getting Data with APIs

A workshop for the curious and data-hungry

go.gwu.edu/apiworkshop

Dan Kerchner kerchner@gwu.edu
Tina Plottel plottel@gwu.edu
Megan Potterbusch mpotterbusch@gwu.edu

Agenda

1. API Basics
2. Exploring the FEC.gov API
3. Using Insomnia client to access the API
4. Demo of NY Times API
5. Ways to use APIs in real life
    - Command-line
    - Python
6. Tips and more APIs

API = Application Programming Interface

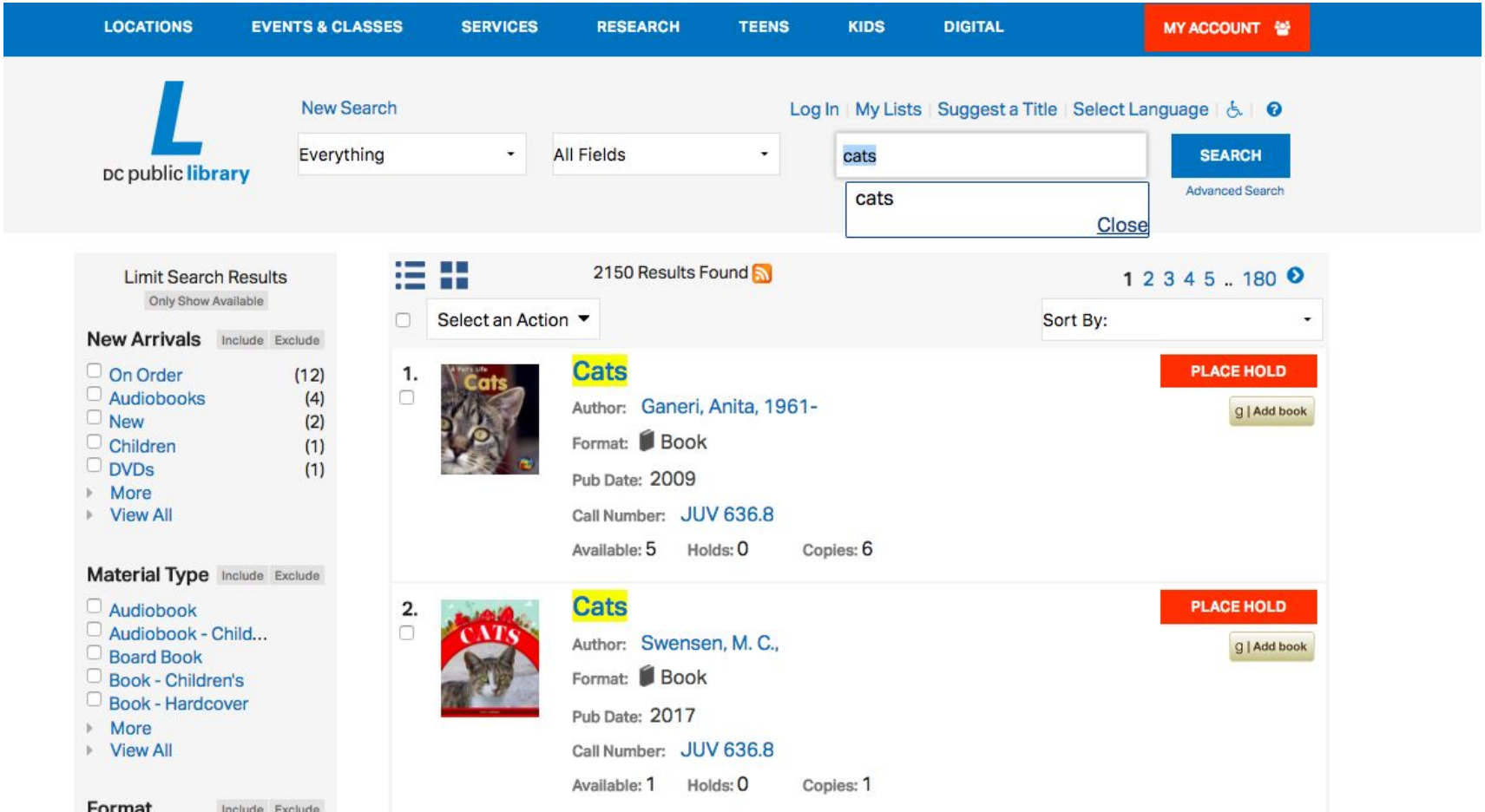
But really, what's an API?

- A web server--**at a URL**--that an application or program can **query** for data.
- Some APIs can be receive data **updates**, too.
- **Syntax** for making requests for data.
- Responses are **structured data**.
- **Response** is formatted for computers (although humans can read it too!)

Why use an API?

- More reliable than “scraping” websites meant for humans to read. Predictable results format that can be retrieved again later.
- You can request just the subset you need. Data may not be available as a downloadable file.
- Gives you structured data you can analyze with code or and statistical software.

Examples of API uses



Our pick

Bodum Brazil

The best French press

This simple but elegant French press makes coffee just as grit-free as brews from competitors that cost much more.

$16* from Amazon

$20 from Target

*At the time of publishing, the price was $20.

The **Bodum Brazil** produced a clear-flavored and consistent cup of coffee that our tasters liked just as much as coffee brewed in most of the other presses we tested, all of which cost more. Its simple glass beaker, steel filtering screens, and plastic exterior appear refined rather than cheap, and its filter pushes down with more ease than other presses' halting, skittish plungers. And you just can't beat the price.

Examples of API uses

For creating websites with dynamic content:

- Showing latest comparison prices on shopping websites
- Embedding maps on a website
- Getting latest stock quotes

For getting data:

- Government statistics
- Social media content
- Citations and publications

Terminology

RESTful or REST APIs
Can be simply accessed at URLs, using HTTP.

HTTP: protocol for requests and responses on the web. Your browser uses this. REST APIs do too!

- URL
- Method (GET, POST)
- List of Headers
- Body

Response formats

- JSON
- XML
- CSV
- RDF
- and more

JSON: JavaScript Object Notation

`{ key: value, key: value, … }`

keys are strings

a value may be a:

- string - in quotes: `"language"`
- number
- boolean: `true` or `false`
- another JSON object
- an array `[ ]` of values
- `null`

JSON: Example

keys are all strings

```
{
    'name': 'George Washington',
    'id': 123456,
    'is_president': true,
    'horses_names': ['Blueskin', 'Nelson'],
    'social_media_ids':
        {'Twitter': '@ICannotTellALie',
         'Github': 'MountVernonCoder'},
    'gw_parking_decal_number': null
}
```

value is a string

value is a number

value is a boolean

value is a list of values

value is another JSON object

value is a `null`

More Terminology

Base URL or endpoint
First part of the URL, sometimes includes a path
https://api.usa.gov/crime/fbi/ucr/estimates/states/*{stateabbrev}*

Parameters
Values that go in the endpoint or query string.
http://myapi.com/animals/search?**q=kittens**&**color=black**

API Key
Unique string of letters and numbers included in each request. Used to make sure you aren't exceeding rate limits (e.g. 1,000 calls/hour).

Let's try using an API!

Federal Elections Commission (FEC)

Campaign Finance Data

www.fec.gov/data/

And some "data journalism" that used this FEC API!

https://www.cnn.com/2018/10/25/politics/party-fundraising-september-zip-code-analysis/index.html

Lots to explore here... look for bulk data

Browse full advanced data sets

Raising | Spending | Candidates | Committees

Filings and reports | Loans and debts | Bulk data and other sources | External data sources

Get started with campaign finance data

Raising

This graph shows how much candidates, party committees and political action committees (PACs) have reported raising, up to specific points in time. Although the graph displays these numbers month-by-month, different committee types have different reporting schedules.

CUMULATIVE AMOUNT RAISED BY COMMITTEES

Explore data | Methodology

FEC Campaign Finance Data

Bulk CSV datasets are available for **download**.

Good for looking at activity in the aggregate.

Requires filtering dataset to get at the precise dataset you need (e.g. individual contributions to a candidate's campaign).

Download bulk data

Downloadable bulk data files contain data from statements and reports filed with the Commission in a form that may be useful to users performing in-depth campaign finance research. The files, which were previously located on the Commission's file transfer protocol (FTP) server, can be very large because they contain transaction-level data. The update schedule of these files varies from daily to weekly. Expand each file's accordion to read more about what is contained in the file.

- All candidates
- Candidate master
- Candidate-committee linkages
- House/Senate current campaigns
- Committee master
- PAC summary
- Contributions by individuals
- Contributions to candidates
- Any transaction from one committee to another
- Operating expenditures
- Electronically filed reports (.fec files)
- Paper filed reports (.fec files)

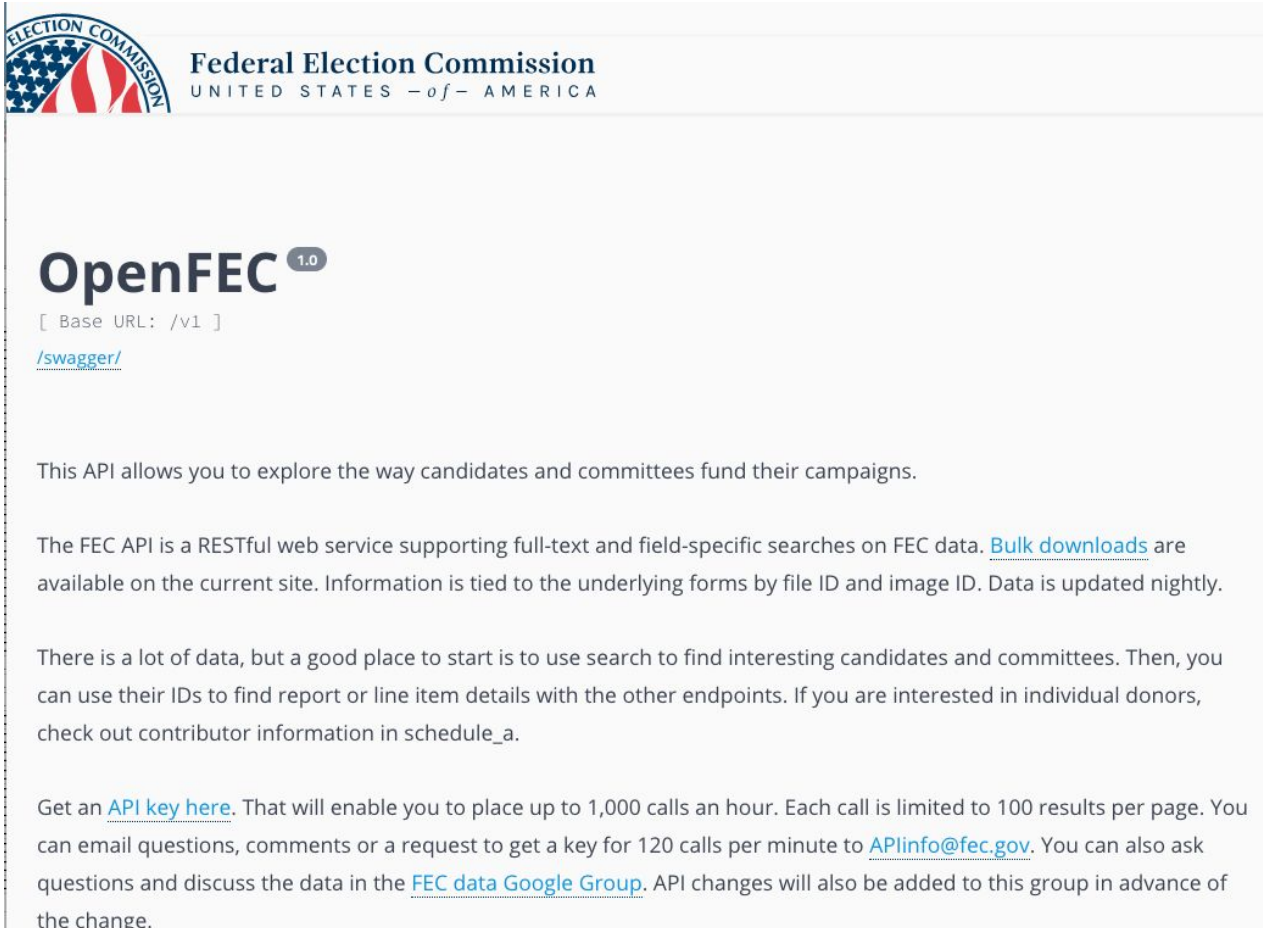
More ways to explore data

Find elections by location

Search candidate or committee profiles

Use the OpenFEC API

https://api.open.fec.gov/developers



Federal Election Commission
UNITED STATES —of— AMERICA

OpenFEC 1.0

[Base URL: /v1]

/swagger/

This API allows you to explore the way candidates and committees fund their campaigns.

The FEC API is a RESTful web service supporting full-text and field-specific searches on FEC data. Bulk downloads are available on the current site. Information is tied to the underlying forms by file ID and image ID. Data is updated nightly.

There is a lot of data, but a good place to start is to use search to find interesting candidates and committees. Then, you can use their IDs to find report or line item details with the other endpoints. If you are interested in individual donors, check out contributor information in schedule_a.

Get an API key here. That will enable you to place up to 1,000 calls an hour. Each call is limited to 100 results per page. You can email questions, comments or a request to get a key for 120 calls per minute to APIinfo@fec.gov. You can also ask questions and discuss the data in the FEC data Google Group. API changes will also be added to this group in advance of the change.

Making API requests with Insomnia

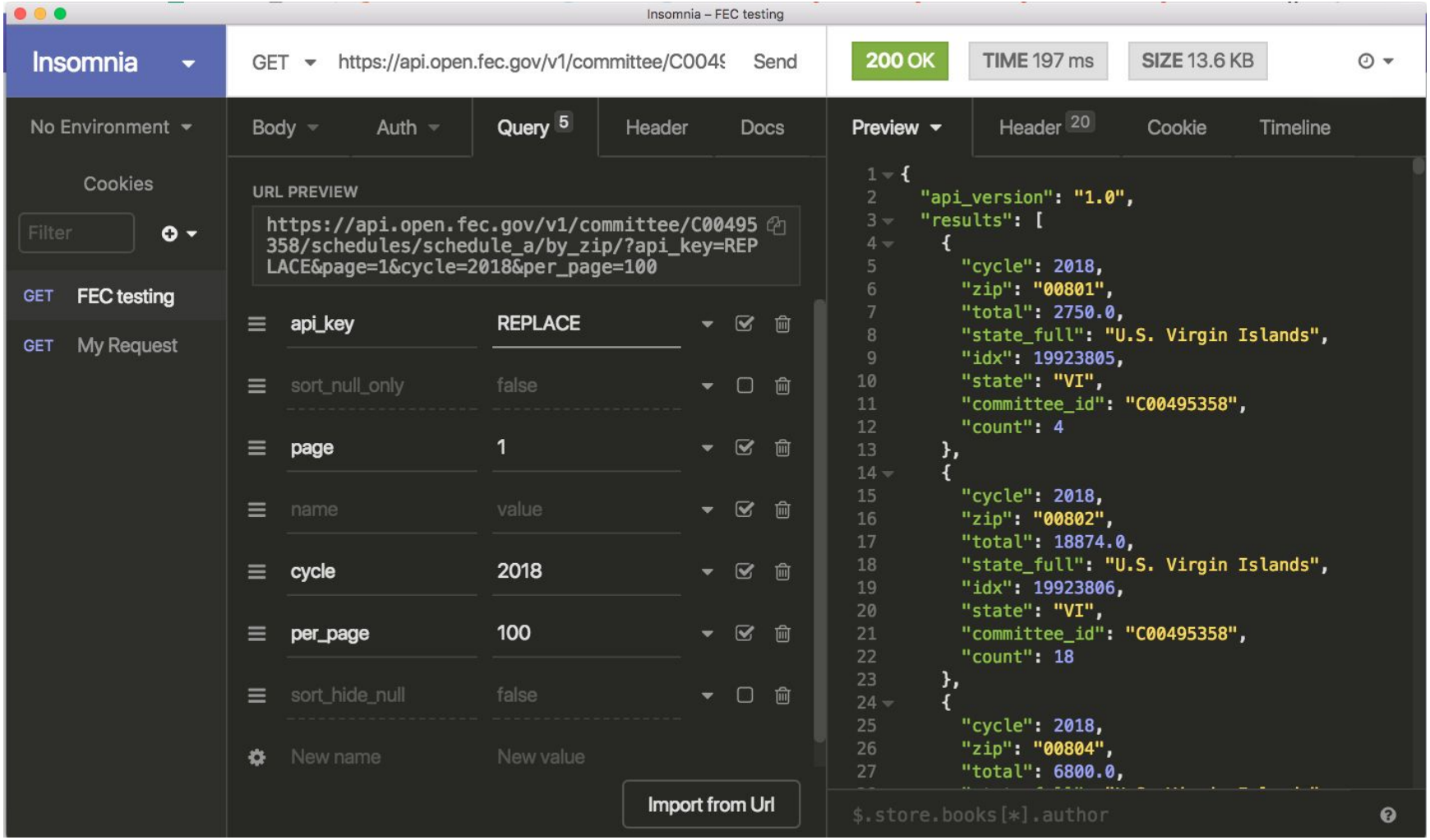
Small application for constructing API calls.

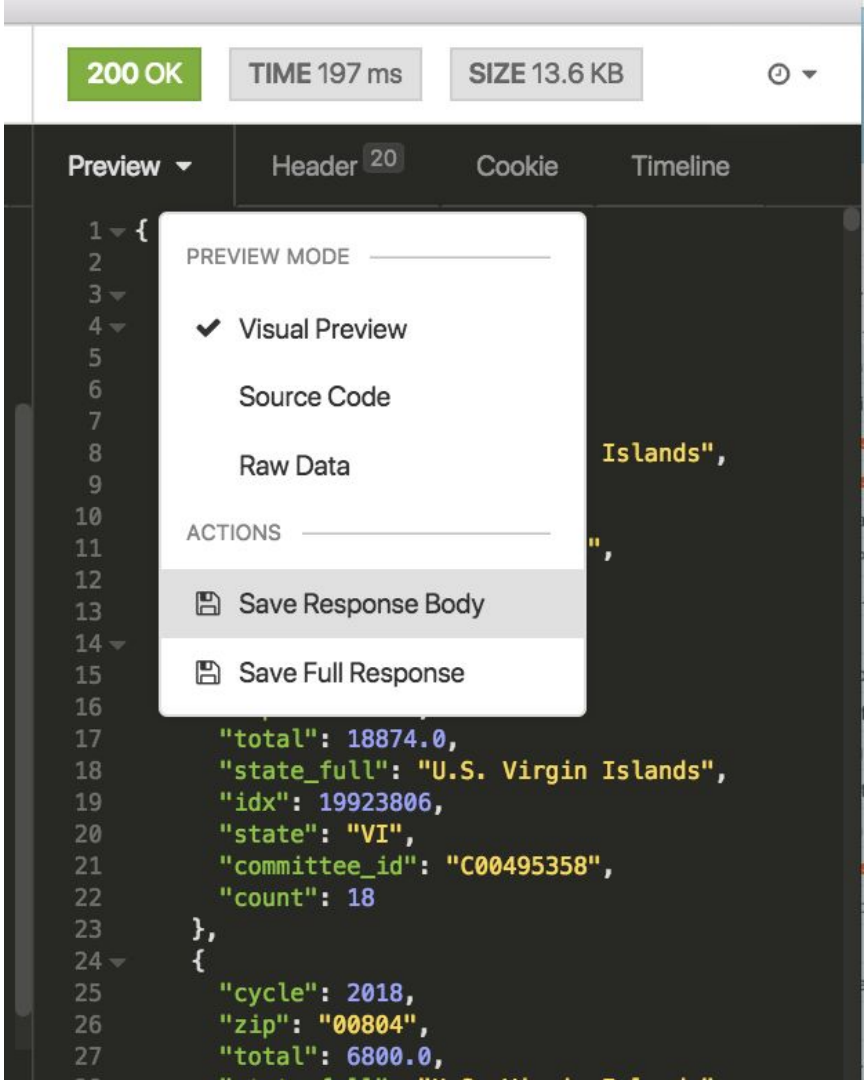
Useful for trying things out, probably not how you'd do it for real.

Getting set up:

Request an API Key:
api.data.gov/signup/

Download and install Insomnia REST client
insomnia.rest


Insomnia – FEC testing
Insomnia
GET
https://api.open.fec.gov/v1/committee/C004
Send
200 OK
TIME 197 ms
SIZE 13.6 KB
No Environment
Cookies
Filter
GET FEC testing
GET My Request
Body
Auth
Query 5
Header
Docs
URL PREVIEW
https://api.open.fec.gov/v1/committee/C00495358/schedules/schedule_a/by_zip/?api_key=REPLACE&page=1&cycle=2018&per_page=100
api_key
REPLACE
sort_null_only
false
page
1
name
value
cycle
2018
per_page
100
sort_hide_null
false
New name
New value
Import from Url
Preview
Header 20
Cookie
Timeline
{
"api_version": "1.0",
"results": [
{
"cycle": 2018,
"zip": "00801",
"total": 2750.0,
"state_full": "U.S. Virgin Islands",
"idx": 19923805,
"state": "VI",
"committee_id": "C00495358",
"count": 4
},
{
"cycle": 2018,
"zip": "00802",
"total": 18874.0,
"state_full": "U.S. Virgin Islands",
"idx": 19923806,
"state": "VI",
"committee_id": "C00495358",
"count": 18
},
{
"cycle": 2018,
"zip": "00804",
"total": 6800.0,
$.store.books[*].author


200 OK
TIME 197 ms
SIZE 13.6 KB
Preview
Header 20
Cookie
Timeline
PREVIEW MODE
Visual Preview
Source Code
Raw Data
ACTIONS
Save Response Body
Save Full Response
Islands",
",
"total": 18874.0,
"state_full": "U.S. Virgin Islands",
"idx": 19923806,
"state": "VI",
"committee_id": "C00495358",
"count": 18
},
{
"cycle": 2018,
"zip": "00804",
"total": 6800.0,

FREE

ProPublica Congress API

Source	Various
Date Released	April 2016
Updates	At least daily

VIEW DOCUMENTATION →

The Congress API returns the following types of data:

- **Roll-call vote data:** Only roll-call votes (not voice votes or division votes) are tracked by official Congressional data sources. Along with basic vote data, the ProPublica API returns additional information that is less readily available, such as party totals. Votes are available from 1991 for the House of Representatives and from 1989 for the Senate.
- **Member data:** Along with general biographical information for current and past members of Congress, the API returns data about members' Congressional roles. Role data includes the Congress number and chamber, as well as the member's title, state and party. A single member may have more than one role in a particular Congress (for example, the member may switch parties or move from the

REQUEST AN API KEY

To request an API key, please provide the following information.

Name *required

Email *required

Organization

I intend to use this API for a commercial application

Purpose

developer.nytimes.com

The New York Times Developers

Get NYT API Key

The New York Times Developer Network

All the APIs Fit to POST

You already know that NYTimes.com is an unparalleled source of news and information. But now it's a premier source of data, too — why just read the news when you can hack it?

Getting Started

The Times Developer Network is our API clearinghouse and community. Here's how to get started:

1. Request an API key
2. Read the API documentation, FAQ and Terms of Use
3. Use the API Tool associated with each API to experiment without writing code

Using The NYT APIs

Considerations

Terms of use

- OK to use for commercial purposes?
- Guidance about how you can display data to the public
- Platform-specific privacy concerns
- Citation

Rate Limits

- Limits number of requests per time period

Pagination

Accessing APIs in real life

- Web application
- Command line
- Code

Web application

go.gwu.edu/sfm

Command line

```
$ curl -X GET
"https://api.open.fec.gov/v1/candidates/?sort_null_only=false&sort=
name&per_page=20&page=1&sort_hide_null=false&api_key=DEMO_KEY" -H
"accept: application/json" | jq
```


```
{
  "api_version": "1.0",
  "pagination": {
    "pages": 1969,
    "per_page": 20,
    "count": 39372,
    "page": 1
  },
  "results": [
    {
      "has_raised_funds": false,
      "party": "IND",
      "first_file_date": "2002-01-30",
      "election_districts": [
        "00"
      ],
      "incumbent_challenge": "C",
      "candidate_id": "P40002172",
```

Code (Python example)

```
# Prints out a list of all candidate IDs and candidate names, using pagination

import requests

# Substitute below for DEMO_KEY with your key from https://api.data.gov/signup/
key = 'DEMO_KEY'
url = 'https://api.open.fec.gov/v1/candidates/?sort_null_only=false&sort=name&' + \
      'per_page=100&page=1&api_key=' + key
r = requests.get(url).json()
num_pages = r['pagination']['pages']

for page in range(1, num_pages + 1):
   url = 'https://api.open.fec.gov/v1/candidates/?sort_null_only=false&sort=name' + \
         '&per_page=100&page=' + str(page) + '&api_key=' + key
   r = requests.get(url).json()
   for candidate in r['results']:
       print(candidate['candidate_id'] + ', ' + candidate['name'])
```

Jupyter notebook

Chronicling America API

Historic newspapers back to 1690 available at the Library of Congress via an API.

https://github.com/LibraryOfCongress/data-exploration/

APIs you might be interested in

- Data USA API
- US Census API
- Bureau of Labor Statistics API
- Folger Shakespeare Library
- Chronicling America
- Metropolitan Museum of Art
- Case Law Project API
- API Directory

Interested in finding more APIs?

Workshop on Finding & Assessing Data

October 30, 3:30-5:00pm

Room 219

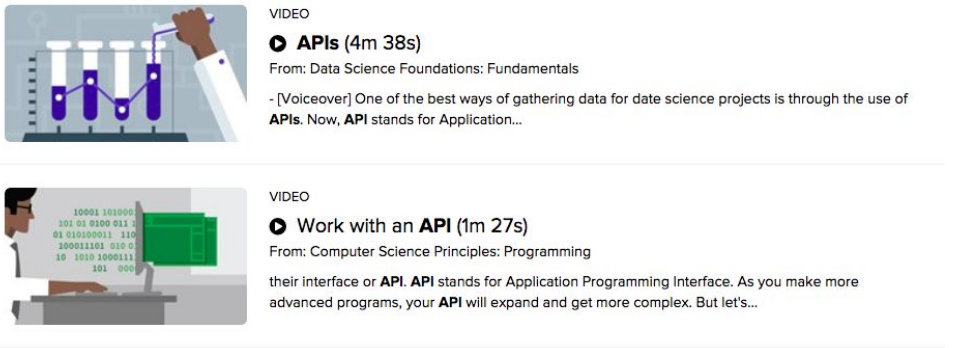
Learn More

[Introduction to APIs (from Zapier)]

[What is an API? In English, please.]

[Lynda.com videos]

(lynda.it.gwu.edu)



Questions? Need help?

Dan Kerchner kerchner@gwu.edu
Tina Plottel plottel@gwu.edu
Megan Potterbusch mpotterbusch@gwu.edu

go.gwu.edu/apiworkshop

Upcoming workshops on Python and R:
Python: Nov 12 R: Nov 15

Coding consultations: calendly.com/gwul-coding