Getting Data with APIs

A workshop for the curious and data-hungry

Laura Wrubel lwrubel@gwu.edu
Dan Kerchner kerchner@gwu.edu
Tina Plottel plottel@gwu.edu

Agenda

- 1. API Basics and the Crime Data API (15 mins)
- 2. Using Hurl.it to access the API (30 mins)
- 3. Demo of NY Times API
- 4. Ways to use APIs (15 mins)
 - Command-line
 - Python
- 5. Tips and more APIs (10 mins)

API = Application Programming Interface

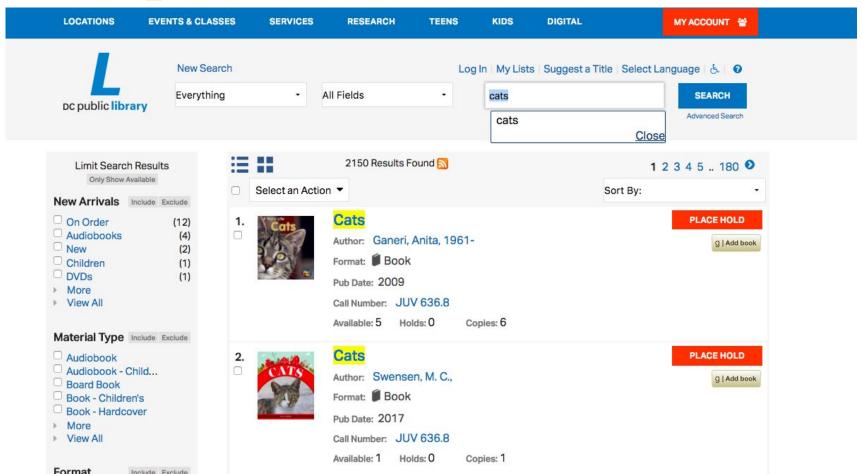
But really, what's an API?

- 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







The <u>Bodum Brazil</u> 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.

Search Search All sustainability

Articles

Sustainability Wei, L, Mei, Y, Chang, CC, DiGiovanni, K WATER ENVIRONMENT RESEARCH, 2016

Sustainability You, SH, Zhang, G, Yang, XH, Chang, CC, Di ...

WATER ENVIRONMENT RESEARCH, 2015 Sustainability

Jones, R.T Journal of the Southern African Institute of Mining

and Metallurgy, 2016 → See all 752.687 articles

Databases

No matching database titles or descriptions found. Please browse Databases by subject.

Books & Media

Sustainability Thiele, Leslie Paul Polity Press 2016.

GW and other WRLC Libraries Sustainability

Pfister, Thomas, (Professor of sociology);... Routledge 2016. Other WRLC Libraries

Sustainability Portney, Kent E London. England 2015. GW and other WRLC Libraries

→ See all 63.468 items

Journals

SUSTAINABLE DEVELOPMENT

SUSTAINABILITY

Library Website

GW ScholarSpace Deposit Form

Sustaining Sustainability in the Trump Years: A Forum on How We Can Preserve Environmental Progress ... GW students are invited to attend Sustaining

Sustainability to share their concerns and to get a better sense of where ...

... Studies Institute for Space Policy Institute for Sustainability Research, Education and Policy Integrated Information, ... Services Office of Student Opportunities Office of Sustainability Office of Technology Transfer Office of The General

Presidents of George Washington University ... and leading its transformation into a model of urban sustainability. Since his inauguration he has worked to generate new ...

→ See all matches

Research Guides

Publishing in the Sciences Debbie Bezanson

Research: From selecting a topic to writing the bibliography

→ See all 138 journals

→ Next 5 Journals

SUSTAINABLE FACILITY

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

Let's try using an API!

An official website of the United States government

FEDERAL BUREAU OF INVESTIGATION

Crime Data Explorer

This site is under development, and will be updated periodically. Crime statistics data is subject to change.

Home

Explorer

Downloads & Documentation

About

Improving access to crime data

The Crime Data Explorer makes nationwide crime data accessible to a wide range of users. <u>View trends</u>, <u>download bulk datasets</u>, and access the <u>Crime Data API</u> for reported crime at the national, state, and agency levels.

Explore by location and type of crime

Novada

All violent crime

View results

FBI Crime Data Explorer

https://crime-data-explorer.fr.cloud.gov

- Visual tool to browse crime data
- Downloads for datasets
- API Explorer
- Documentation about API endpoints and parameters

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. (e.g. a state or year) 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).

Requesting an API Key

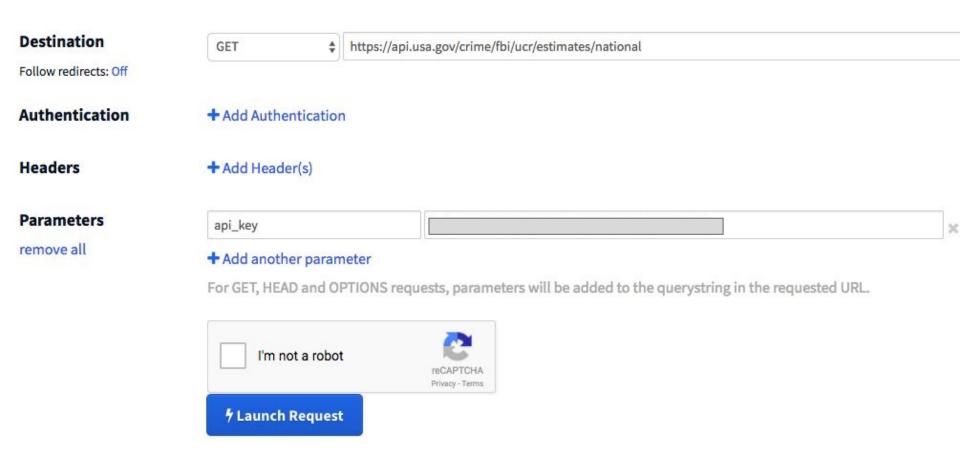
https://api.data.gov/signup/

Making API requests

https://www.hurl.it

Website for constructing API calls.

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



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',
                                                 value is a string
 'id': 123456,
                                                 value is a number
 'is president': true,
                                                 value is a boolean
 'horses names': ['Blueskin', 'Nelson'], value is a list of values
 'social media ids':
                                                 value is another JSON object
     {'Twitter': '@ICannotTellALie',
      'Github': 'MountVernonCoder'},
                                                 value is a null
 'gw parking decal number': null
```



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

information	1.		
Name *re	quired		
			A
Email *rec	guired		
Organiza	tion		
		 ,	mmercial

https://developer.nytimes.org



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

Rate Limits

limits number of requests per time period

Pagination

Accessing APIs in real life

- Web application
- Command line
- Code

Web application

http://go.gwu.edu/sfm



Command line

\$ python twarc.py timeline nytimes | jq

```
"contributors": null,
"truncated": false,
"text": "=This tweet begins with an equal sign.",
"is_quote_status": false,
"in_reply_to_status_id": null,
"id": 810920620867407900,
```

Twarc command-line tool for collecting from Twitter https://github.com/docnow/twarc

Code

```
import tweepy
auth = tweepy.OAuthHandler(consumer key, consumer secret)
auth.set access token(access token, access token secret)
api = tweepy.API(auth)
public_tweets = api.home_timeline()
for tweet in public tweets:
   print tweet.text
```

https://github.com/tweepy/tweepy

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

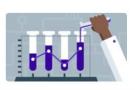
- Data USA API
- US Census API
- Bureau of Labor Statistics API
- Folger Shakespeare Library
- Chronicling America
- Metropolitan Museum of Art
- API Directory

Learn More

Introduction to APIs (from Zapier)

What is an API? In English, please.

<u>Lynda.com videos</u>



VIDEC

O APIs (4m 38s)

From: Data Science Foundations: Fundamentals

- [Voiceover] One of the best ways of gathering data for date science projects is through the use of APIs. Now, API stands for Application...



VIDEO

O Work with an API (1m 27s)

From: Computer Science Principles: Programming

their interface or API. API stands for Application Programming Interface. As you make more advanced programs, your API will expand and get more complex. But let's...

Questions? Need help?

Laura Wrubel lwrubel@gwu.edu
Dan Kerchner kerchner@gwu.edu
Tina Plottel plottel@gwu.edu



Need help with something? Get in touch. Make a Research Consultation appointment or a

Programming/Coding Consultation appointment

Credits: This workshop was inspired by the <u>DPLA</u> "Introduction to APIs" workshop and draws on their slides for explaining APIs.