

# Getting Data with APIs



A workshop for the curious and data-hungry

Laura Wrubel	<a href="mailto:lwrubel@gwu.edu">lwrubel@gwu.edu</a>
Dan Kerchner	<a href="mailto:kerchner@gwu.edu">kerchner@gwu.edu</a>
Tina Plottel	<a href="mailto:plottel@gwu.edu">plottel@gwu.edu</a>

# 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

LOCATIONS

EVENTS & CLASSES


SERVICES


RESEARCH

TEENS



KIDS



DIGITAL

MY ACCOUNT 



New Search

Log In | My Lists | Suggest a Title | Select Language |  

Everything  All Fields 

cats

cats

Close

SEARCH

Advanced Search

Limit Search Results

Only Show Available

New Arrivals

Include Exclude

☐ On Order

(12)

☐ Audiobooks

(4)

☐ New

(2)

☐ Children

(1)

☐ DVDs

(1)

More

View All

Material Type

Include Exclude

☐ Audiobook

☐ Audiobook - Child...

☐ Board Book

☐ Book - Children's



☐ Book - Hardcover


More


View All


Format


Include Exclude

2150 Results Found 


1 2 3 4 5 .. 180 

Select an Action 

Sort By: 


1.

☐



Cats

Author: Ganeri, Anita, 1961-

Format:  Book

Pub Date: 2009

Call Number: JUV 636.8


Available: 5   Holds: 0   Copies: 6

PLACE HOLD

g | Add book


2.

☐



Cats

Author: Swensen, M. C.,

Format:  Book

Pub Date: 2017

Call Number: JUV 636.8

Available: 1   Holds: 0   Copies: 1

PLACE HOLD

g | Add book

## 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.

Search All

sustainability

Search

## 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

SUSTAINABILITY

SUSTAINABLE DEVELOPMENT

SUSTAINABLE FACILITY

→ [Next 5 Journals](#)

→ [See all 138 journals](#)

## Library Website

[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 ...

[GW ScholarSpace Deposit Form](#)

... 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](#)



# 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. [View trends](#), [download bulk datasets](#), and access the [Crime Data API](#) for reported crime at the national, state, and agency levels.

## Explore by location and type of crime

Nevada



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?](http://myapi.com/animals/search?q=kittens&color=black)**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.

## Destination

Follow redirects: [Off](#)

GET



https://api.usa.gov/crime/fbi/ucr/estimates/national

## Authentication

[+ Add Authentication](#)

## Headers

[+ Add Header\(s\)](#)

## Parameters

[remove all](#)

api\_key



[+ Add another parameter](#)

For GET, HEAD and OPTIONS requests, parameters will be added to the querystring in the requested URL.



I'm not a robot



reCAPTCHA  
[Privacy](#) - [Terms](#)

**Launch Request**



# 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'},

'gw\_parking\_decal\_number': null

value is a null

}

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

# 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>

[Social Feed Manager](#) [Collection Sets](#) [Credentials](#) [Exports](#) [Monitor](#)

[Collection Sets](#) / [115th U.S. Congress](#) / U.S. Senators (115th Congress) Official Twitter User Timelines

## U.S. Senators (115th Congress) Official Twitter User Timelines

 **Twitter user timeline**  
Collection is active. Turn off to edit.

[Turn off](#)  
[Deactivate](#)  
[Edit](#)  
[Export](#)

Next harvest at Nov. 6, 2017, 12:28:07 p.m. EST

**Description:** Twitter user timelines belonging to U.S. Senators in the 115th Congress. List comes from the Senate website, archived here: <https://web.archive.org/web/20170127080714/https://www.senate.gov/senators/contact/>. Individual handles were mostly taken from links on each Senator's website. When there was no handle linked on the Senator's website, their name was searched for on Twitter. In all such cases, the handle was Verified. For more information, see: [https://docs.google.com/a/email.gwu.edu/document/d/1OILAwskM5U7ePAsYIKbBhjhdKEmKsQVl\\_V6SgCjluQ/edit?usp=sharing](https://docs.google.com/a/email.gwu.edu/document/d/1OILAwskM5U7ePAsYIKbBhjhdKEmKsQVl_V6SgCjluQ/edit?usp=sharing)

**Data collected:** 44 files (133.9 MB)

**Stats:**

- tweets: 374,922

[Details](#)

Seeds

[Download seed list](#)



# 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.](#)

[Lynda.com videos](#)



VIDEO

▶ **APIs** (4m 38s)

From: Data Science Foundations: Fundamentals

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



VIDEO

▶ **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



Ask Us

Need help with  
something? Get  
in touch.

Make a [Research Consultation appointment](#)  
or a  
[Programming/Coding Consultation appointment](#)