

# 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 (accessed at a URL) that another application or program can query for data.
- Can sometimes receive data/updates, too.
- Communication is formatted for computers, not humans.
- Has a syntax for making **requests** for data.
- Has a defined structure for **responses**.

# Why use an API?

- To get data using software or code.
- Data may not be available as a downloadable file, or be too big. You can get just what you need.
- Updating data within an application.
- Polite way to get data from a website: better than “scraping” it off web pages meant for human users.
- Gives you structured data you can use for other purposes, including analysis.

# Examples of API uses


## The Washington Post

*Democracy Dies in Darkness*

### doom by floodwaters

The house floats on the Baker River in Warren, N.H. And then comes a bridge.


Taylor Turner · 21 hours ago



### Two Argentine towns were bombarded with hail storms, one like a cannon, the other like a shotgun

One town got pummeled with softball-size stones; the other buried under five feet of pellets. Here's how that happened.

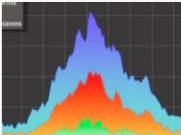
Jason Samenow · 23 hours ago




### Hurricane season usually winds down in November, but this hasn't been a normal season

The odds favor at least one more storm this season.

Brian McNoldy · 23 hours ago



### Trick or treat! Here's the hour-by-hour weather forecast for



#### National Airport

Dulles BWI

**Right Now**

**55.0 °**  
(12.8° C)

**Weather** Overcast

**Wind** North at 5.8 mph


**Wind Chill** 54 F


**Dew Point** 43.0 F

**Pressure** 1025.2 mb

Last updated: Nov 1 2017, 11:52 a.m.

Tweets by @capitalweather

 **Capital Weather Gang** @capitalweather  
Standing out in the crowd: Here are 7 of the nation's coolest TV meteorologists: [wapo.st/2A6IGSV](http://wapo.st/2A6IGSV)



**Perspective | Standing out in the crowd: Here are 7 of L...**  
See if your hometown favorite made the list.  
[washingtonpost.com](http://washingtonpost.com)

34m

Capital Weather Gang Retweeted

# Examples of API uses

- Confirming your address when making online orders
- Posting articles or websites to social media
- Embedding maps on a website
- Getting latest stock quotes



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](#)



# Terminology

## RESTful or REST APIs

Can be accessed at URLs with directory paths, uses HTTP requests explicitly.

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

- URL
- Method
- List of Headers
- Body

## Methods

- **GET:** requests data from the server
- **POST:** requests that data be changed on the server
- **DELETE:** requests that data be deleted

# 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. For example, 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

<http://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

jqCeuyGkbINUxnMXTTyqRyMqXsJQHKUtW8vtHLI



[+ 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': 'Justin Littman',
  'id': 123456,
  'is_librarian': true,
  'children_names': ['Sam', 'Maya'],
  'social_media_ids':
    {'Twitter': '@justin_littman',
     'Github': 'justinlittman'},
  '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

# 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/1OILAwskM5U7ePAsYIKbBjhjhdKEmKsQVl\\_V6SgCjluQ/edit?usp=sharing](https://docs.google.com/a/email.gwu.edu/document/d/1OILAwskM5U7ePAsYIKbBjhjhdKEmKsQVl_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

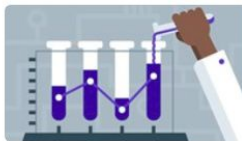
- [Data USA API](#)
- [US Census API](#)
- [Bureau of Labor Statistics API](#)
- [Folger Shakespeare Library](#)
- [Chronicling America](#)
- [Metropolitan Museum of Art](#)
- [API Directory](#)

# 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