

# How to find data

---

What are you even looking for?

---

## Anything that is...

- Inspected (restaurants, amusement parks, buildings)
- Licensed (doctors, taxis, pets, hunting)
- Enforced (speeding tickets, subway stops)
- Purchased (contracts, Medicare claims data, etc.)
- Reported (311 calls, etc.)

... likely lives in a database.

# Knowing what to look for

- Your need to figure out:
  - What kind of data might be helpful to your reporting project?
  - What data like that exists?
  - Who has it?
  - What is it called?
  - How can you get it?
- To answer those questions: **REPORT!**
  - Find and call sources
    - experts/academics/researchers
    - other reporters
    - bureaucrats
    - people who work in the field (lawyers, teachers, business people, etc)
  - The internet (more on this)

# Does a dataset exist?

Evidence that a dataset exists:

- A form (digital or paper), with a specific title and/or form number
- An online lookup tool
  - [NYC Physician discipline](#)
- Published reports
- Research by academics

## Get it by...

- Googling
- Asking the government
- Asking other organizations
- DIY: surveys, crowdsourcing, scraping

# Googling for data

---

# Google

- There are a bunch of advanced search functions that can help you find data or documents
- Search specific phrases with quotes “”
  - E.G. “new york population”
- Exclude search words with a hyphen -
  - E.G. “new york population” -city
- Either/or search with OR
  - E.G. budget OR appropriations
- Find websites that are similar to others by using related
  - E.G. related:propublica.org

# Google

- Search only on a certain domain
  - E.G. site:propublica.org; site:nyc.gov
- Search for a specific file type
  - E.G. filetype:pdf; E.G. site:nyc.gov filetype:xls budget
  - Try these handy file types
    - ppt (PowerPoint)
    - xls (Excel pre-2007)
    - xlsx (Excel post-2007)
      - CHECK FOR BOTH xls and xlsx!
    - doc (Word document)
    - txt (text files)
    - csv (comma-delimited text files, usually to store data)

# Google

You can put all these search techniques together!

- Find data or documents from specific agencies
  - E.G. site:nyc.gov filetype:xls budget
- Refine your search with the OR operator
  - E.G. site:nyc.gov filetype:xls subway OR budget

There's more! [https://www.google.com/advanced\\_search](https://www.google.com/advanced_search)

# GIVE ME ALL THE DATA



Table Preview View Data Create Visualization

Uniq...	Cre...	Close...	Agen...	Agen...	Com...	Desc...	Loca...	Incid...	Incid...	Stree...
4330033	07/17/20...		DOT	Departm...	Traffic Sig...	Controller		11217		
43295363	07/17/20...		DOT	Departm...	Traffic Sig...	Controller				
43295426	07/17/20...		DOT	Departm...	Street Lig...	Street Lig...		11225	305 CRD...	CROWN S...
43299570	07/17/20...		DOT	Departm...	Traffic Sig...	Pedestria...		11234		
43294785	07/17/20...		DOT	Departm...	Street Lig...	Street Lig...		11412		
43297680	07/17/20...		DOT	Departm...	Traffic Sig...	Controller		10306		
43297311	07/17/20...		DOT	Departm...	Street Co...	Pothole		11215	357 3 STR...	3 STREET
43293252	07/16/20...		DOT	Departm...	Street Lig...	Street Lig...			628 5 ST E	5 ST E
43301216	07/16/20...	07/17/20...	DOT	Departm...	Traffic Sig...	Controller		10473		
43293242	07/16/20...		DOT	Departm...	Street Lig...	Street Lig...			839 220 S...	220 ST E
43293243	07/16/20...		DOT	Departm...	Street Lig...	Lamppos...		11226	1600 CAT...	CATON A...
43296515	07/16/20...		DOT	Departm...	Street Lig...	Traffic Sig...		10306		
43300898	07/16/20...		DOT	Departm...	Street Lig...	Traffic Sig...		10306		
43299244	07/16/20...		DOT	Departm...	Street Lig...	Street Lig...		11435		

< Previous   Next >   Showing Rows 1-14 out of 21,148,752



**NYC OpenData**

## 311 Service Requests from 2010 to Pres

Social Services

View Data Visualize Export ...

All 311 Service Requests from 2010 to Present. This information is auton...

Updated July 18, 2019   Data |

About this Dataset

# GIVE ME ALL THE DATA

Many government websites have interactive tools to display data

- E.G. <https://opendata.cityofnewyork.us/>
- Socrata is a popular vendor
- These often have limited functionality (sorting/filtering)
- Instead, find the option to download the data
- Look for a Download or Export option.
- Try different formats, e.g. text file, csv, excel, pdf

# Government data + FOIA

---

# Government Data

## Step 1: call the agency

- Never send a records request cold
- Make friends, be nice
- Describe what you're looking for, or what you're trying to find
  - Learning the name that the agency uses to describe the database is GOLD
- Start by just asking. Who knows, they might say yes
  - Formal requests take time
- Don't bank on a records request – but do file EARLY.

# Government Data

- Federal: Freedom of Information Act (FOIA)
- State: each state has their own
  - New York State's is called Freedom of Information Law (FOIL)
  - <https://www.dos.ny.gov/coog/index.html>
  - Learn the response deadlines for your state
  - What's allowable in one state might not be in another
  - States often have First Amendment coalitions/advocacy groups. Find yours.
- Privacy considerations

# FOIA/FOIL tips

- Before filing, make sure you're requesting from the right agency. This will save you time
- Include the specific name and description of the data/documents in your request
- Ask for electronic fulfillment (unless for some reason you specifically want paper)

# FOIA/FOIL language to include

Format: we prefer to receive records in the following formats, listed in order of preference:

- (1) an electronic data format such as a spreadsheet, delimited data set, database file, or similar;
- (2) other non-proprietary electronic format;
- (3) word processing file, text-based PDF, or similar;
- (4) paper copies.

Please also provide any and all documentation related to such electronic records, including but not limited to data dictionaries, database documentation, record layouts, code sheets, data entry instructions, and similar printed or electronic documentation materials.

## FOIA/FOIL tips, cont.

- Ask for the request to be fulfilled on a “rolling basis,” i.e. release the documents as they come
- Agencies have deadlines for how long they can take to respond. Set a reminder for yourself so you know to follow up. Or keep track with a website like [FOIA Machine](#)

# FOIA/FOIL tips, cont.

## FEES!!

- Ask for a fee waiver, then negotiate fees
- Agencies might overcharge as a denial/stalling tactic
- Per-page reproduction fee? Don't need it if it's electronic
- "It will take \$10,000 for 50 hours to fulfill this request..."
  - How much is that employee paid an hour?

# FOIA/FOIL tips, cont.

There's a balance to strike between being too broad and too narrow

- Too specific and you might inadvertently exclude some records
- Too vague and you run the risk of the agency saying the request is too vast to fulfill
- This is where your previous phone calls come in handy
- Limit time periods. 5 years of records is easier to fulfill than 50

# FOIA/FOIL tips, cont.

- You don't just want the data, you want the stuff that goes with it
- Resources
  - Letter generator: <http://www.splc.org/page/lettergenerator>
  - MuckRock: <https://www.muckrock.com/>
- Know your rights
  - Denied? You can (and should) appeal
  - Be polite, but don't take "no" for an answer
  - Public shaming

