

Getting Open Data through the Western
Pennsylvania Regional Data Center or:

How I Learned to Stop Downloading
CSVs and Love the API

David Walker Steve Saylor Twitter Bot



With special guest:
Digalot Sirius Mischief

What is “Open Data”?

What is the WPRDC?



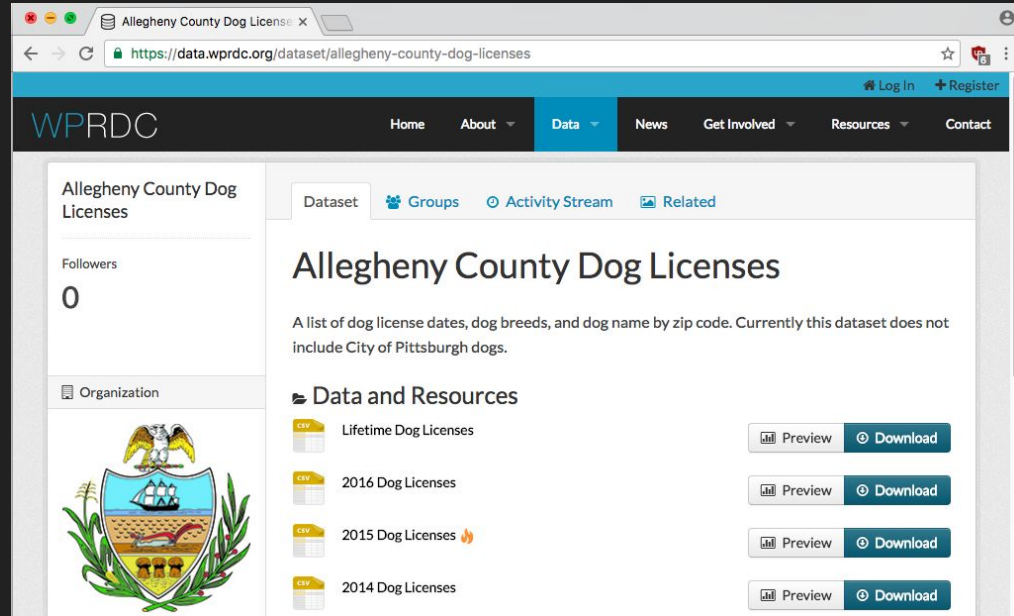
Data? What kind of data?

Civic data. Megatons of it! (155 datasets currently.)

- Updated roughly hourly
 - 311 Requests
 - Air Quality Measurements
 - Snow Plow Tracking
- Updated daily
 - Public Safety (crime blotter, arrests, citations)
 - PLI (Permits, License, and Inspections) Violations
- Updated monthly
 - Restaurant Inspections
 - Property Assessments
 - Property Sales
 - Accidental Drug Overdoses
- Popular yet somewhat static (updated yearly)
 - Car Crashes (Now covering 5 counties!)
 - Dog Licenses

How do I get that delicious data?


Option 1: Download CSV (comma-separated values) files



The screenshot shows a web browser window displaying the WPRDC dataset page for Allegheny County Dog Licenses. The browser's address bar shows the URL <https://data.wprdc.org/dataset/allegheny-county-dog-licenses>. The WPRDC logo is in the top left, and navigation links for Home, About, Data, News, Get Involved, Resources, and Contact are in the top right. The page title is "Allegheny County Dog Licenses". Below the title, there are tabs for "Dataset", "Groups", "Activity Stream", and "Related". The main content area features the title "Allegheny County Dog Licenses" and a description: "A list of dog license dates, dog breeds, and dog name by zip code. Currently this dataset does not include City of Pittsburgh dogs." Below this is a section titled "Data and Resources" with a list of datasets:

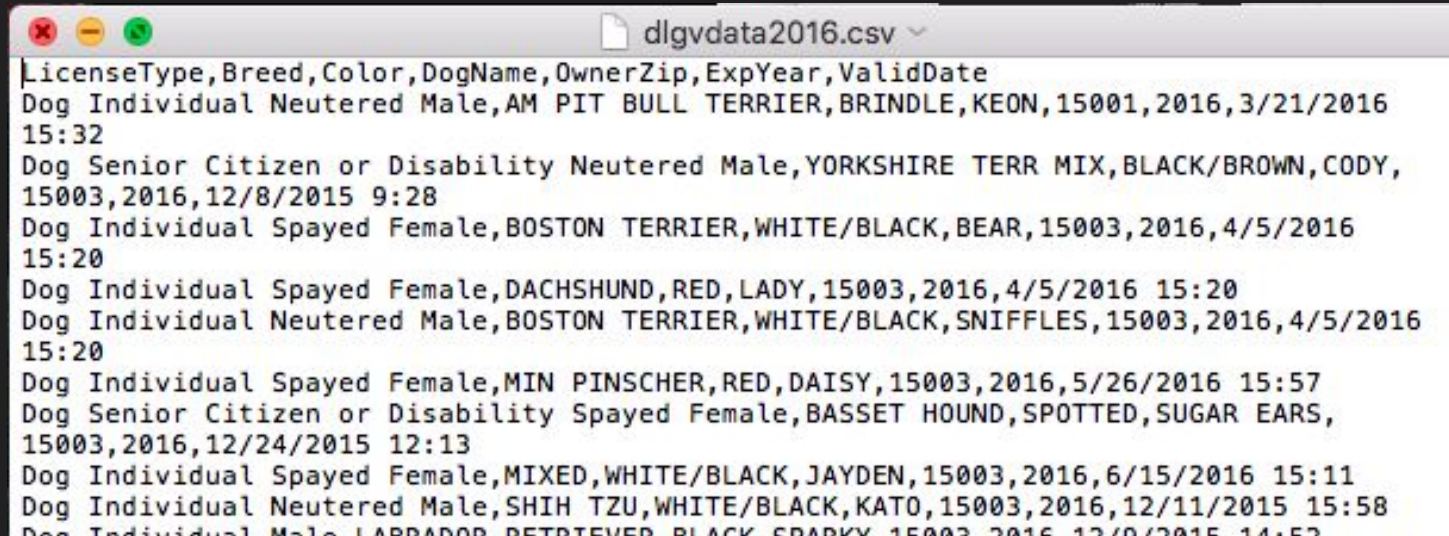
Dataset Name	Preview	Download
Lifetime Dog Licenses	Preview	Download
2016 Dog Licenses	Preview	Download
2015 Dog Licenses 🔥	Preview	Download
2014 Dog Licenses	Preview	Download

On the left side of the page, there is a sidebar with the following information:

- Allegheny County Dog Licenses
- Followers: 0
- Organization: 

How do I get that delicious data?

Option 1: Download CSV (comma-separated values) files



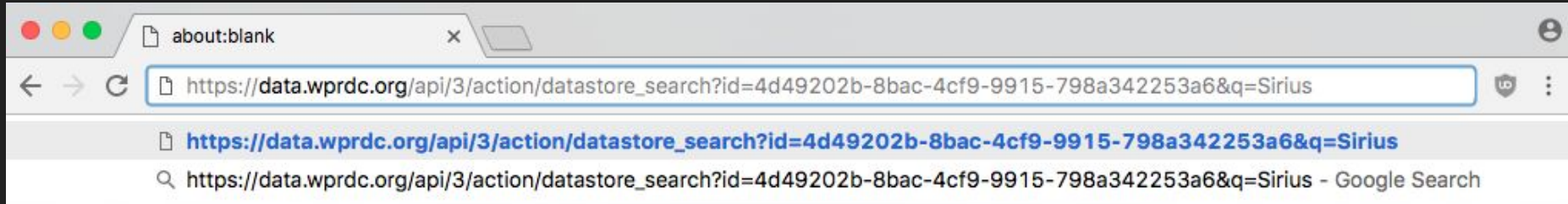
```
dlgvdata2016.csv
LicenseType,Breed,Color,DogName,OwnerZip,ExpYear,ValidDate
Dog Individual Neutered Male,AM PIT BULL TERRIER,BRINDLE,KEON,15001,2016,3/21/2016
15:32
Dog Senior Citizen or Disability Neutered Male,YORKSHIRE TERR MIX,BLACK/BROWN,CODY,
15003,2016,12/8/2015 9:28
Dog Individual Spayed Female,BOSTON TERRIER,WHITE/BLACK,BEAR,15003,2016,4/5/2016
15:20
Dog Individual Spayed Female,DACHSHUND,RED,LADY,15003,2016,4/5/2016 15:20
Dog Individual Neutered Male,BOSTON TERRIER,WHITE/BLACK,SNIFFLES,15003,2016,4/5/2016
15:20
Dog Individual Spayed Female,MIN PINSCHER,RED,DAISY,15003,2016,5/26/2016 15:57
Dog Senior Citizen or Disability Spayed Female,BASSET HOUND,SPOTTED,SUGAR EARS,
15003,2016,12/24/2015 12:13
Dog Individual Spayed Female,MIXED,WHITE/BLACK,JAYDEN,15003,2016,6/15/2016 15:11
Dog Individual Neutered Male,SHIH TZU,WHITE/BLACK,KATO,15003,2016,12/11/2015 15:58
Dog Individual Male,LABRADOR RETRIEVER,BLACK,SPARKY,15003,2016,12/9/2015 14:52
```

How do I get that delicious data?

Option 2: Use our beautiful API (Application Programming Interface)

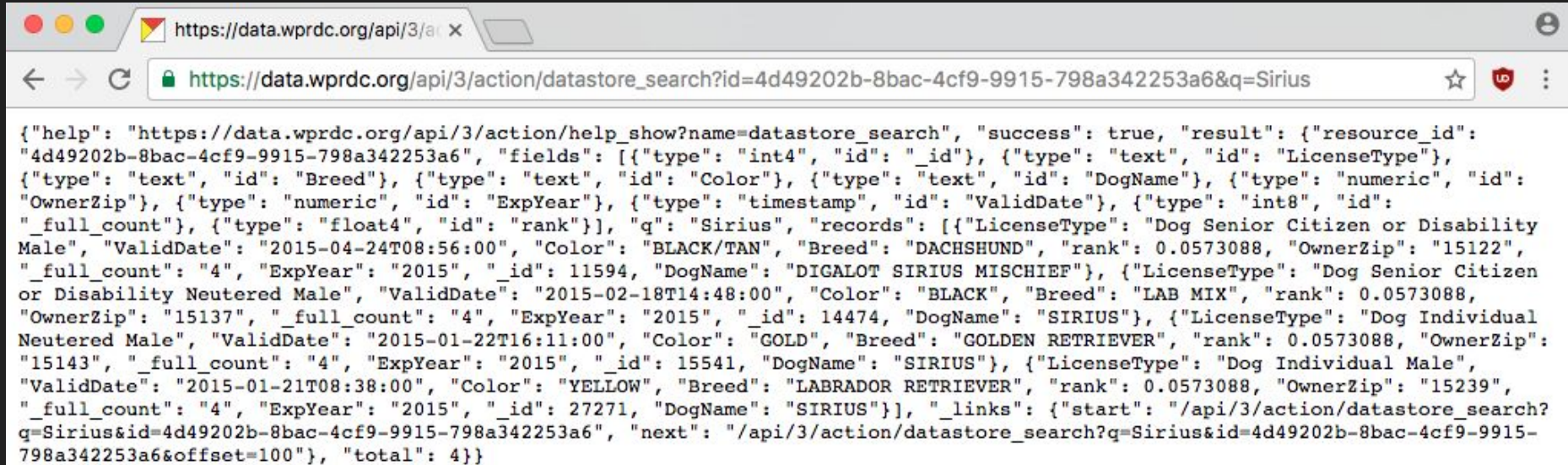
How do I get that delicious data?

Option 2: Use our beautiful API (Application Programming Interface)



How do I get that delicious data?

Option 2: Use our beautiful API (Application Programming Interface)



```
{ "help": "https://data.wprdc.org/api/3/action/help_show?name=datastore_search", "success": true, "result": { "resource_id": "4d49202b-8bac-4cf9-9915-798a342253a6", "fields": [ { "type": "int4", "id": "_id" }, { "type": "text", "id": "LicenseType" }, { "type": "text", "id": "Breed" }, { "type": "text", "id": "Color" }, { "type": "text", "id": "DogName" }, { "type": "numeric", "id": "OwnerZip" }, { "type": "numeric", "id": "ExpYear" }, { "type": "timestamp", "id": "ValidDate" }, { "type": "int8", "id": "_full_count" }, { "type": "float4", "id": "rank" } ], "q": "Sirius", "records": [ { "LicenseType": "Dog Senior Citizen or Disability Male", "ValidDate": "2015-04-24T08:56:00", "Color": "BLACK/TAN", "Breed": "DACHSHUND", "rank": 0.0573088, "OwnerZip": "15122", "_full_count": "4", "ExpYear": "2015", "_id": 11594, "DogName": "DIGALOT SIRIUS MISCHIEF" }, { "LicenseType": "Dog Senior Citizen or Disability Neutered Male", "ValidDate": "2015-02-18T14:48:00", "Color": "BLACK", "Breed": "LAB MIX", "rank": 0.0573088, "OwnerZip": "15137", "_full_count": "4", "ExpYear": "2015", "_id": 14474, "DogName": "SIRIUS" }, { "LicenseType": "Dog Individual Neutered Male", "ValidDate": "2015-01-22T16:11:00", "Color": "GOLD", "Breed": "GOLDEN RETRIEVER", "rank": 0.0573088, "OwnerZip": "15143", "_full_count": "4", "ExpYear": "2015", "_id": 15541, "DogName": "SIRIUS" }, { "LicenseType": "Dog Individual Male", "ValidDate": "2015-01-21T08:38:00", "Color": "YELLOW", "Breed": "LABRADOR RETRIEVER", "rank": 0.0573088, "OwnerZip": "15239", "_full_count": "4", "ExpYear": "2015", "_id": 27271, "DogName": "SIRIUS" } ], "_links": { "start": "/api/3/action/datastore_search?q=Sirius&id=4d49202b-8bac-4cf9-9915-798a342253a6", "next": "/api/3/action/datastore_search?q=Sirius&id=4d49202b-8bac-4cf9-9915-798a342253a6&offset=100" }, "total": 4 } }
```

Mad Libs with Open Data?

Can you explain an example API call?

https://data.wprdc.org/api/3/action/datastore_search?id=1797ead8-8262-41cc-9099-cbc8a161924b&limit=5

Root URL

API URL - URL where all API endpoints are located

API endpoint - the action we want to perform

Question mark - indicates that text afterwards represent parameters

Parameter/value - key-value pair of endpoint-specific parameter mapped to a value we send

Ampersand - separates parameters

Tiny examples, anyone?

Questions?

What can I do with the API?

1) Full-text search:

https://data.wprdc.org/api/action/datastore_search?id=4d49202b-8bac-4cf9-9915-798a342253a6&q=Sirius

What can I do with the API?

2) More powerful queries:

```
https://data.wprdc.org/api/action/datastore_search_sql?sql=SELECT * FROM  
'4d49202b-8bac-4cf9-9915-798a342253a6' WHERE 'DogName' LIKE '%SIRIUS%'
```

These are SQL queries (where SQL is a special language for dealing with databases).

What can I do with the API?

2) More powerful queries:

```
https://data.wprdc.org/api/action/datastore_search_sql?sql=SELECT * FROM  
'4d49202b-8bac-4cf9-9915-798a342253a6' WHERE 'DogName' LIKE '%SIRIUS%'  
AND "Breed" LIKE 'DACHSHUND'
```

These are SQL queries (where SQL is a special language for dealing with databases).

What can I do with the API?

2) More powerful queries:

```
https://data.wprdc.org/api/action/datastore_search_sql?sql=SELECT%20*%20FR  
OM%20%224d49202b-8bac-4cf9-9915-798a342253a6%22%20WHERE%20%22  
DogName%22%20LIKE%20%27%SIRIUS%%27%20AND%20%22Breed%22%2  
0LIKE%20%27DACHSHUND%27
```

These are SQL queries (where SQL is a special language for dealing with databases).

How about that Twitterbot demo?

When should I download CSV files? When should I use the API?

Download CSVs when:

- The data is not going to change.
- You want to manipulate the whole dataset.

Use the API when:

- The data is being regularly updated.
- You don't want to download a 500,000-row dataset just to find the sales prices for one house.
- You want to get the types of the data fields.

What should I do next?

- Some ideas:
 - Open-Data Mad Libs Poetry Slam!
 - Play with the city's Burgh's Eye View (where you can download data)
 - Explore the Tiny Examples notebook
 - Find an interesting dataset and make your own API calls
 - Make your own Twitterbot
 - Build "Even Madder Libs"
 - Create an automatically updating map of the potholes reported in your neighborhood

Is this the final slide?

Yes.

Go to <https://github.com/WPRDC/api-workshop>

for the Tiny Examples notebook, the Twitterbot demo, links to where you can run these notebooks online, and lots of documentation.

If you make something cool, send us a link to it! We'll be blogging about our API talks and would like examples to include in our blog post.

WPRDC site: <http://wprdc.org> / E-mail: wprdc@pitt.edu / Twitter: @wprdc

