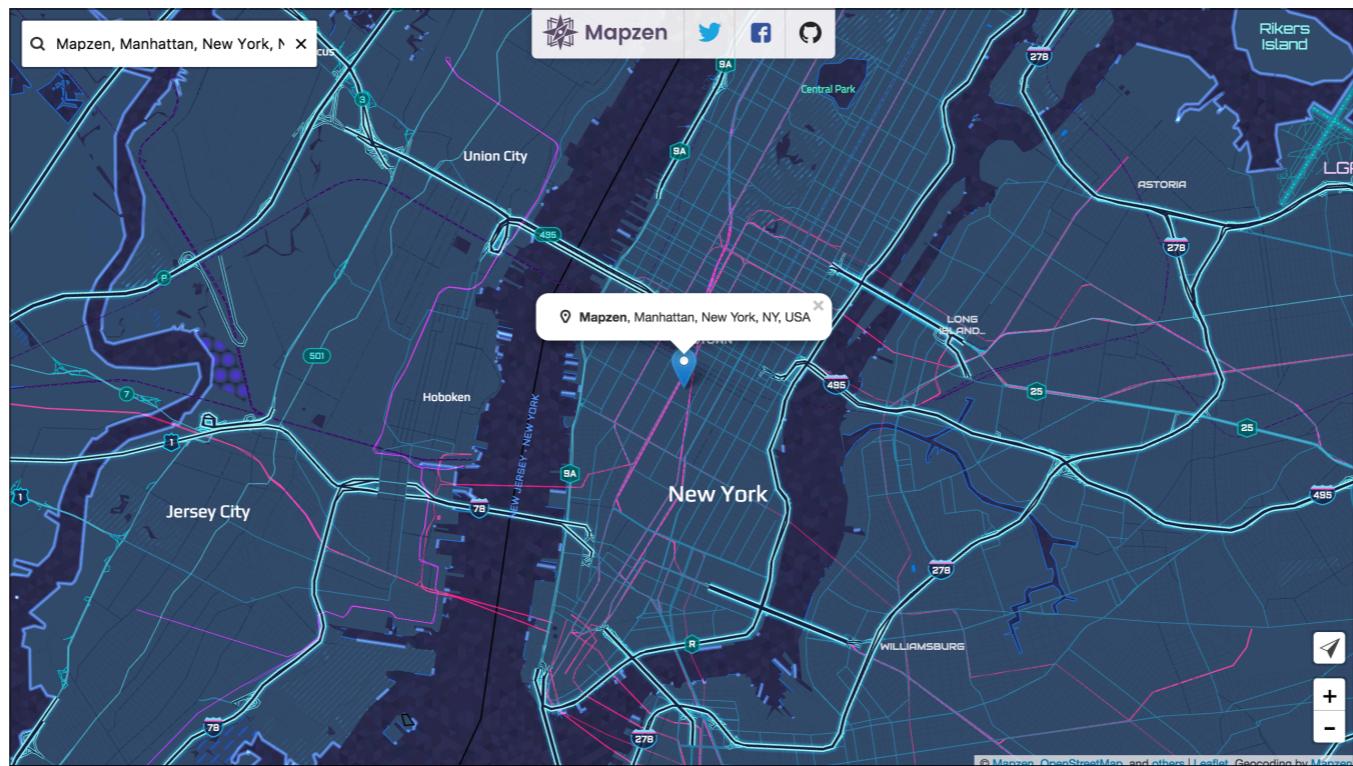


The world is weird and wonderful!

Dan Phiffer @dphiffer
csv,conf,v3



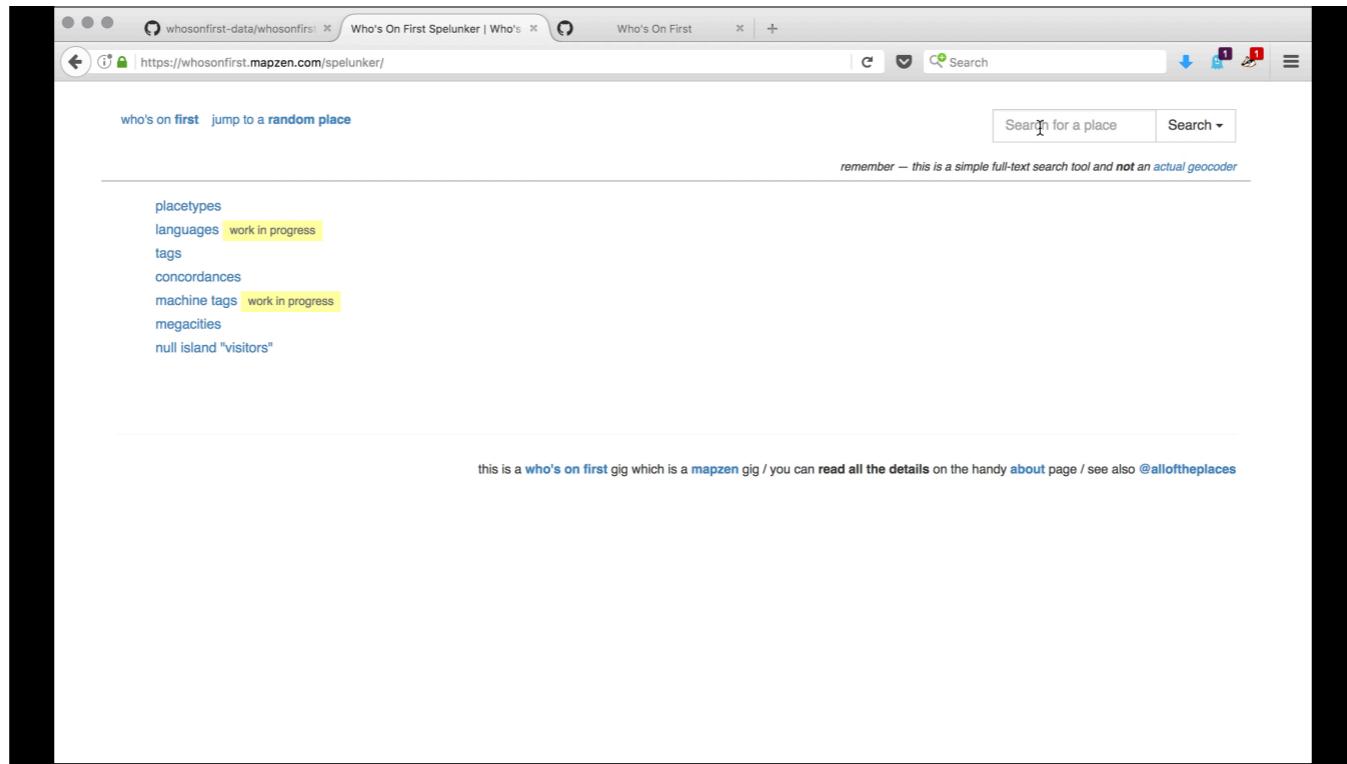
Hi, I am Dan Phiffer and I work on the data team at Mapzen.

We make open source tools to roll your own map.

This talk has two parts: an intro to Who's On First generally, and a demo of some of my recent work on a tool called Boundary Issues.

This may get a little bit technical, but there is a cat video and some dancing.

Also, shout out to my parents who drove down from Washington!



Who's On First aka WOF is a collection of places that lives on GitHub. There are about 26 million records, all open licensed. Each place has a stable numeric ID, kind of like the place's phone number. We use the GeoJSON format to encode details about the place, you can see some different names for Portland here.

places with the name 'Street' #646

 **Closed** missinglink opened this issue on Feb 13 · 4 comments



missinglink commented on Feb 13

Contributor



the following wof records have the name 'Street' which, albeit possible, are unlikely the correct name:

```
Street|default|101852537  
Street|default|404439545  
Street|eng|101852537  
Street|eng|404439545
```

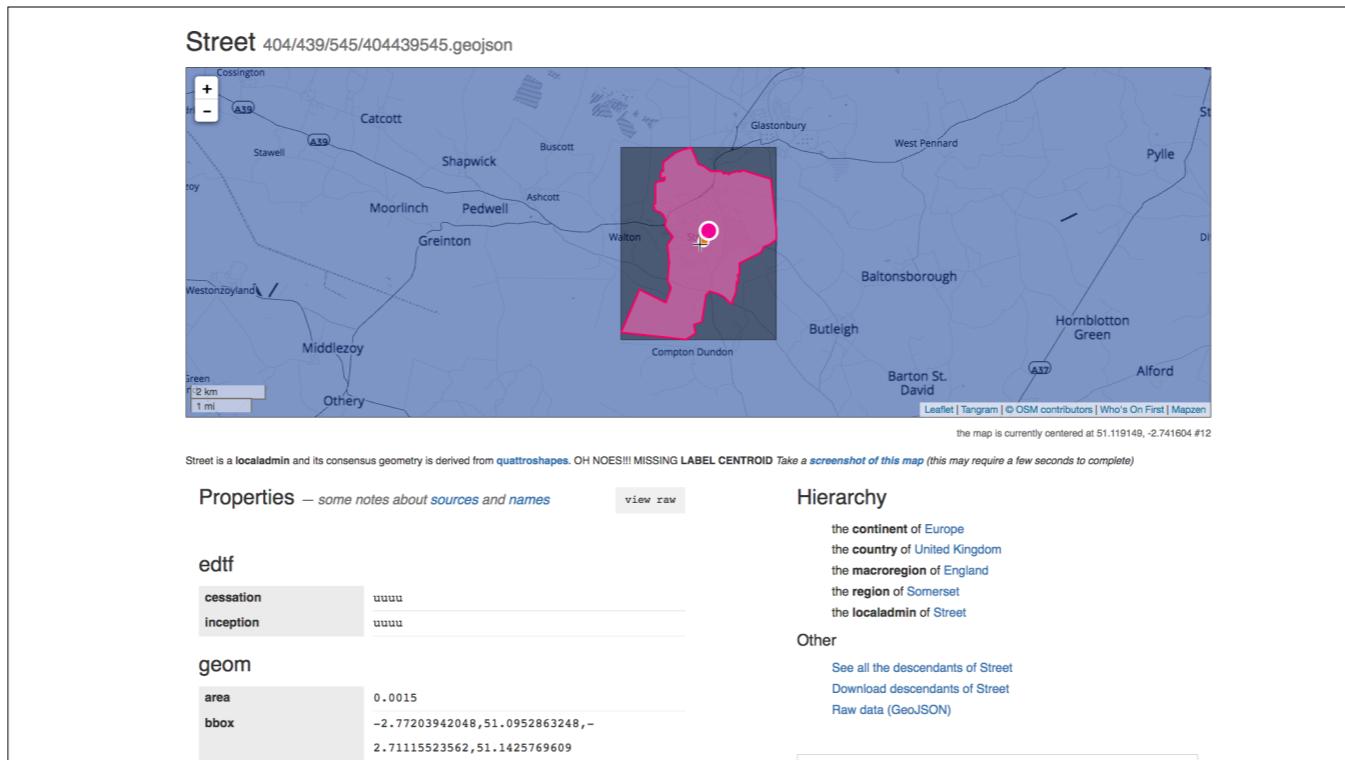


stepps00 self-assigned this on Feb 13

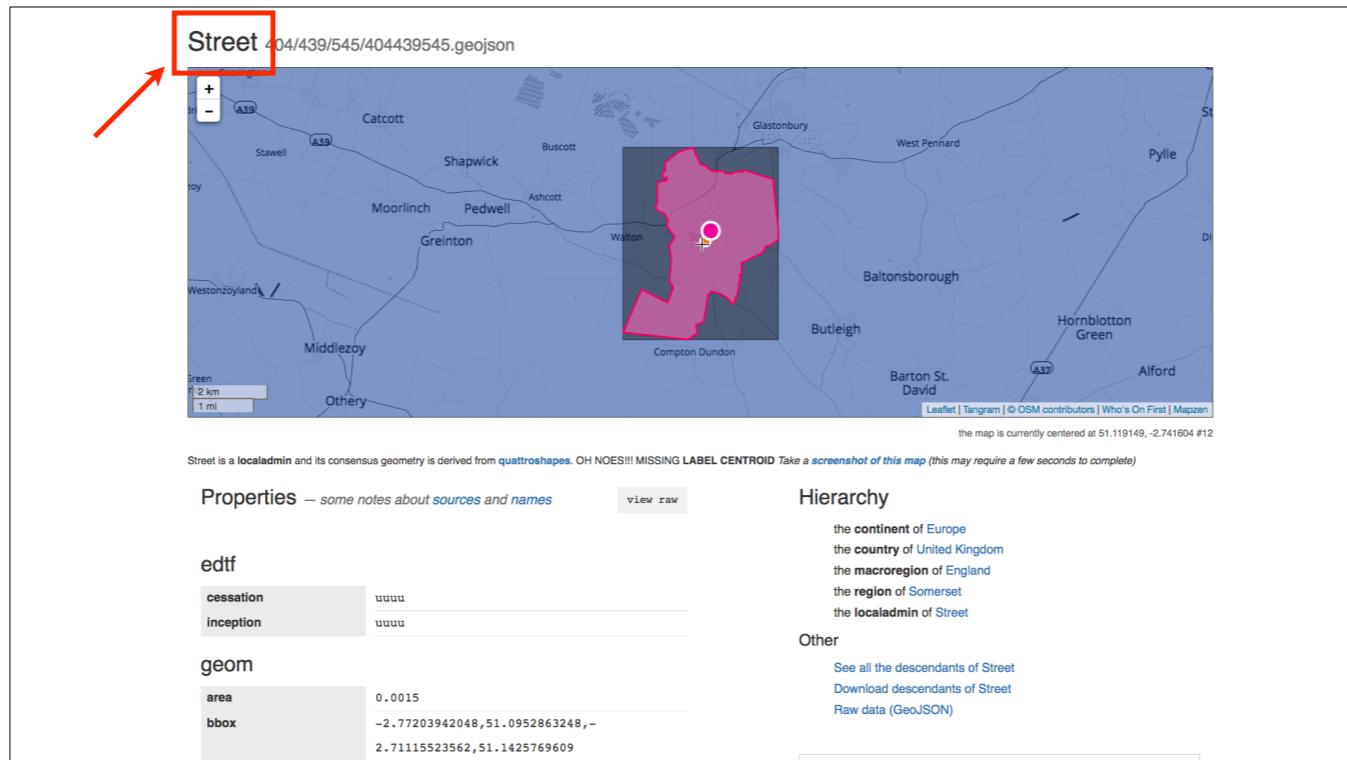


stepps00 added the [enhancement](#) label on Feb 13

One time Peter Johnson (from the Mapzen search team) filed a bug against our data repo. He found some records that looked wrong. Places should not be called “street,” obviously.



As it turns out there are actually places in the world called Street.



This place has very bad SEO.

A screenshot of a GitHub issue comment thread. The top comment is from user **stepps00**, posted on Feb 13. The message reads: "Research of Street, Somerset suggests there is nothing actionable for this issue. The records listed above ([101852537](#) and [404439545](#)) represent a valid, current location in England." Below this, under "Additional references:", are three links: DMOZ, Office for National Statistics, and Somerset Guide. The second comment is from user **stepps00**, also assigned **nvkelso** and unassigned **stepps00**, posted on Feb 13. The message reads: "Over to @nvkelso before I close this issue."

stepps00 commented on Feb 13

Owner +

Research of Street, Somerset suggests there is nothing actionable for this issue. The records listed above ([101852537](#) and [404439545](#)) represent a valid, current location in England.

Street "is a large village and civil parish in the county of Somerset, England." ([source](#)).

Additional references:

- [DMOZ](#)
- [Office for National Statistics](#)
- [Somerset Guide](#)

stepps00 assigned nvkelso and unassigned stepps00 on Feb 13

stepps00 commented on Feb 13

Owner +

Over to @nvkelso before I close this issue.

That the name would be so generic is surprising, but it is also true.

A screenshot of a GitHub issue comment thread. The first comment is from user nvkelso, posted on Feb 13, reading "The world is weird and wonderful! :)" with an "Owner" button and edit/close icons. This comment is closed by nvkelso on Feb 13. The second comment is from user dphiffer, also on Feb 13, reading "Yo @nvkelso I'm gonna steal that as the title for a conference talk proposal." with an "Owner" button and edit/close icons. A single smiley face emoji is shown below this comment.

nvkelso commented on Feb 13

The world is weird and wonderful! :)

Owner + edit X

nvkelso closed this on Feb 13

dphiffer commented on Feb 13

Yo @nvkelso I'm gonna steal that as the title for a conference talk proposal.

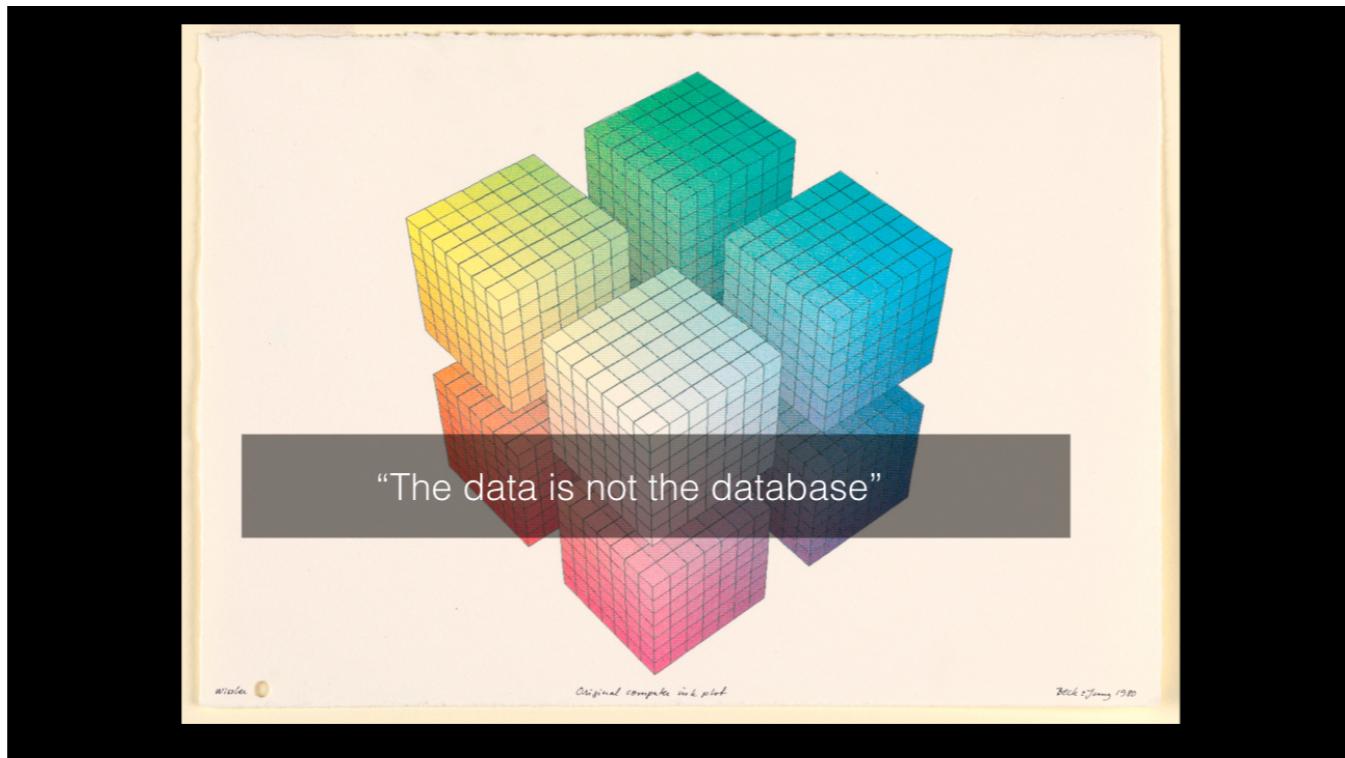
Owner + edit X

1

And that is the story of how this talk got its name.



Who's On First was started in 2015 by Aaron Straup Cope.

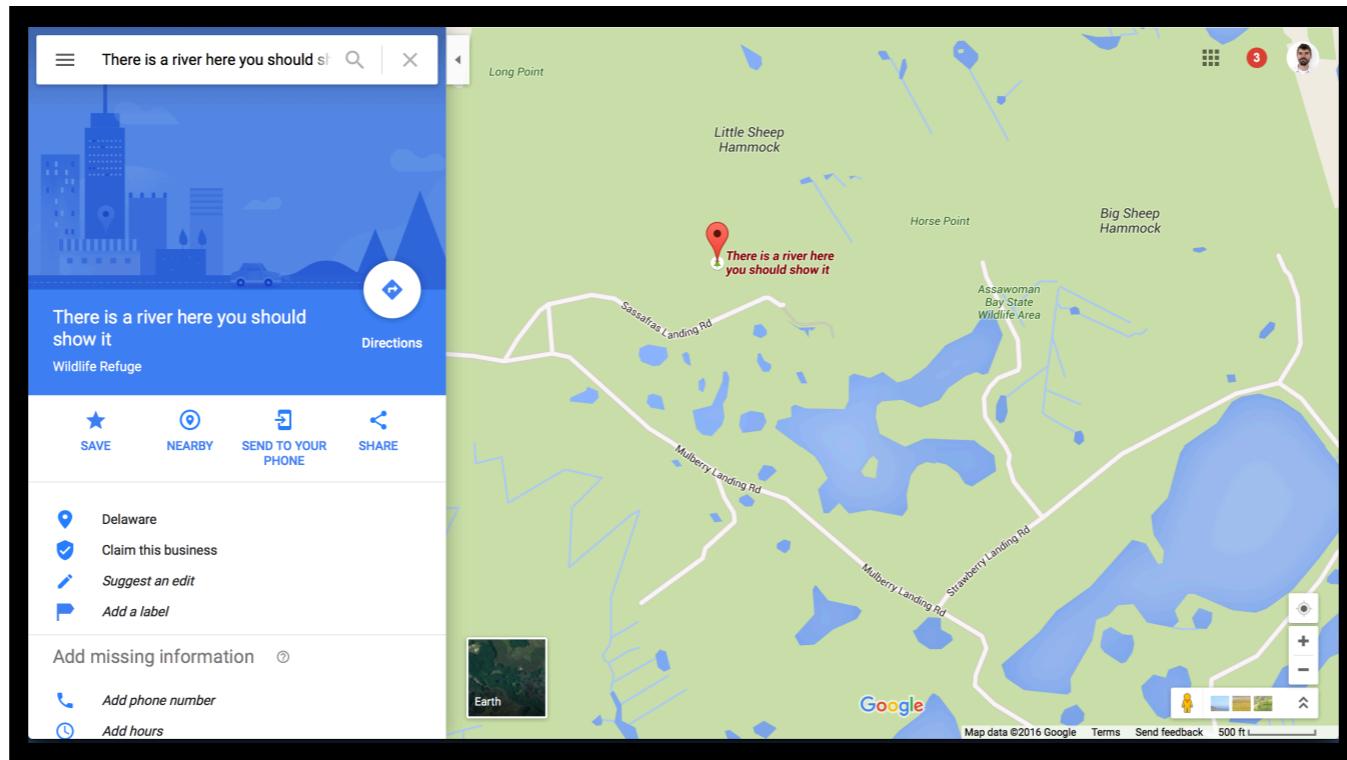


This is from Aaron's slide deck, and it's written up in a blog post called [Mapping With Bias](#).

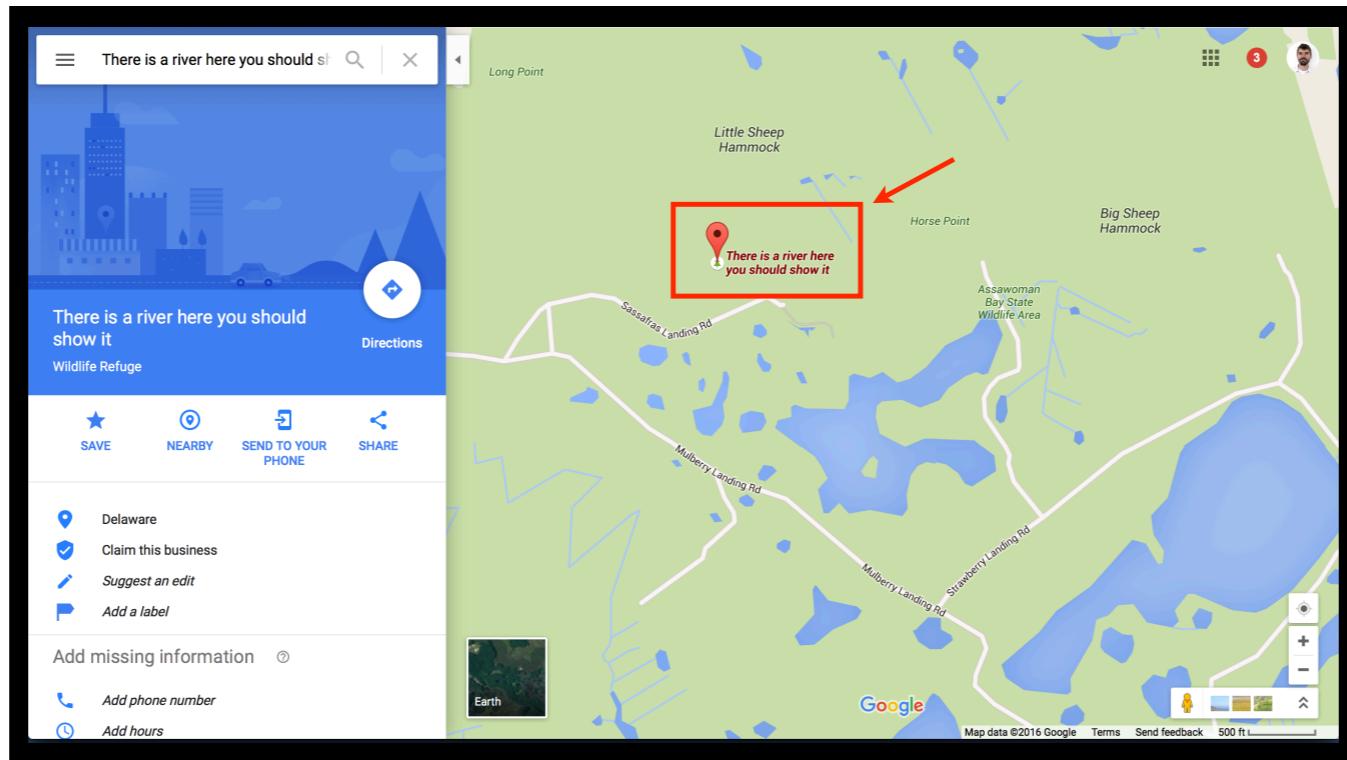
We encode our records as plain text files for their stability, and also because that lets you decide which database platform you prefer.

You like Postgres? Fine. ElasticSearch? Hey we use that too! SQLite, oh yes.

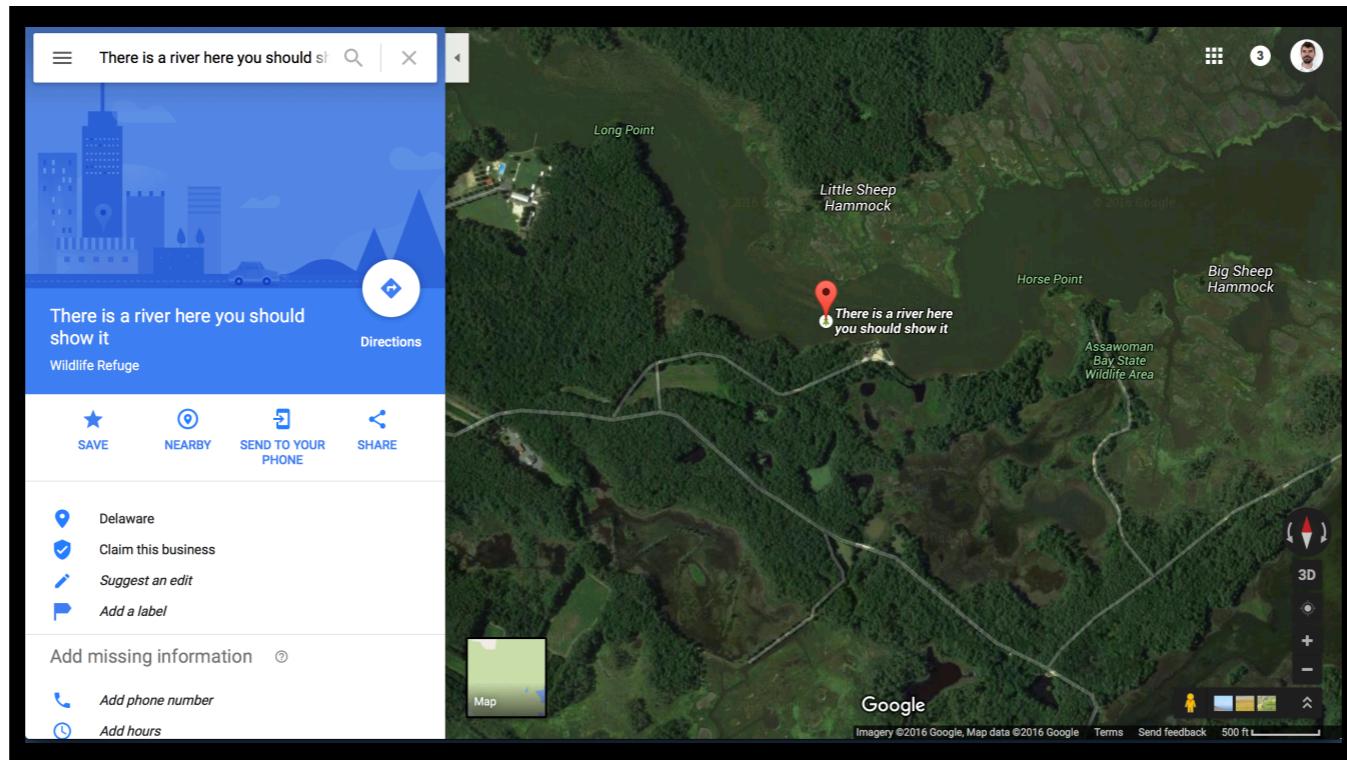
We offer an API, but you can also just download all the records and index them yourself.



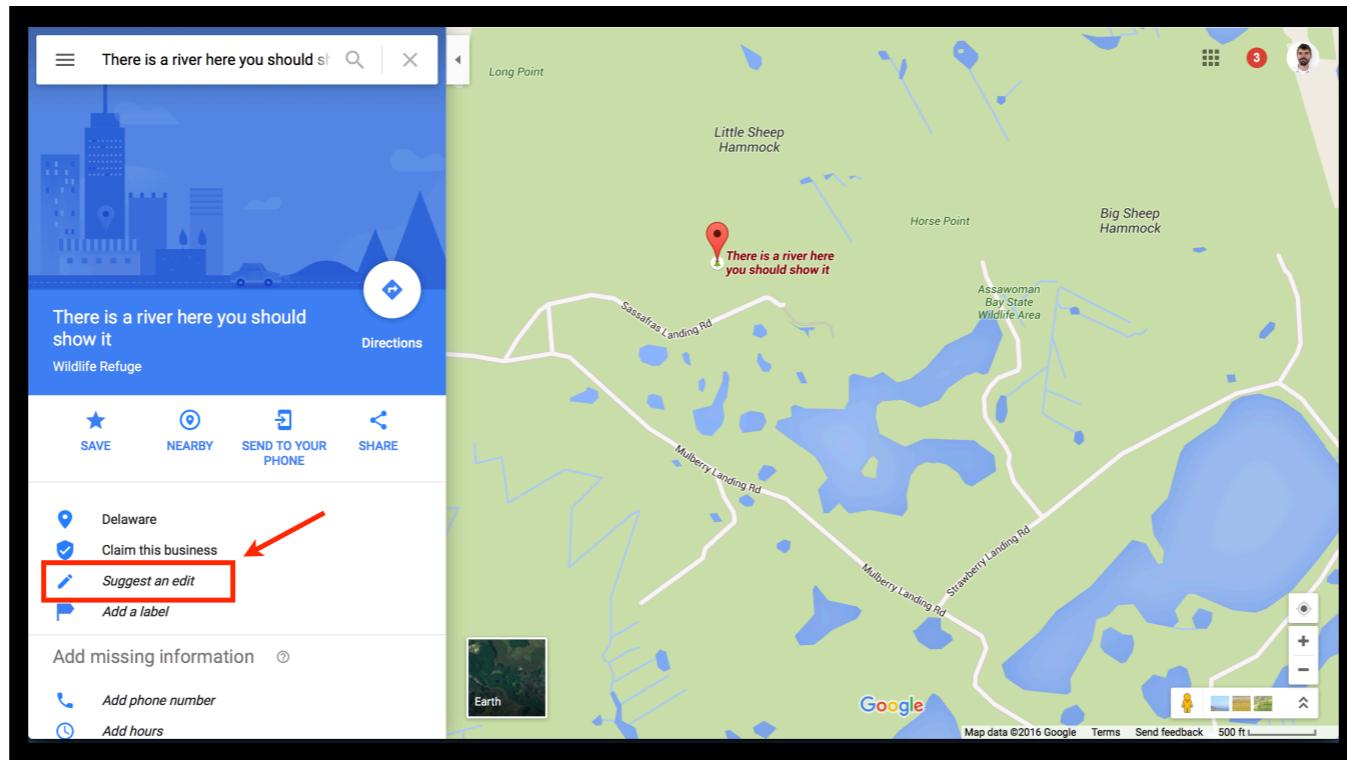
This is **not** a Mapzen map, but it does illustrate some problems we're trying to solve.



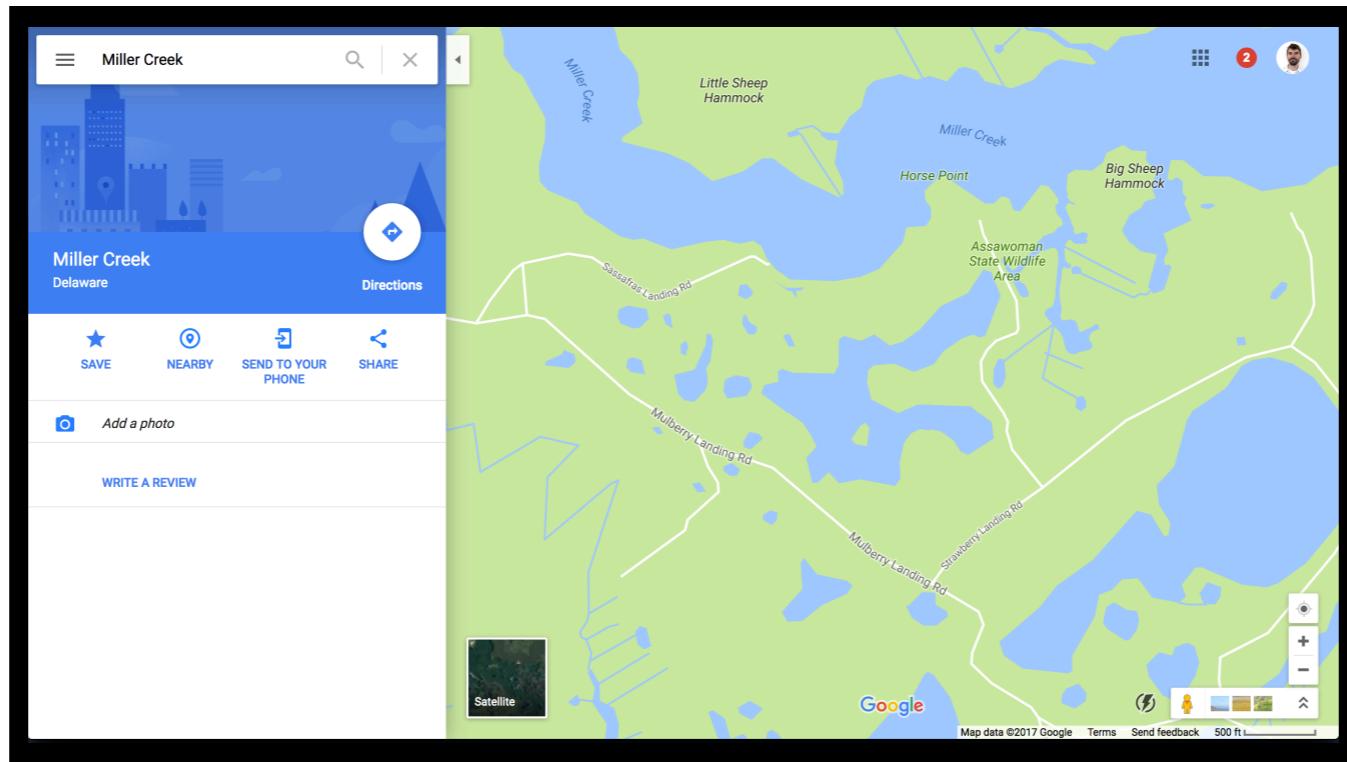
This is a creative use of the tools at hand.



Sure enough, there is a river here.



This is the type of thing I think about. What should happen next when you click on this button?



As of this morning the river had been added.



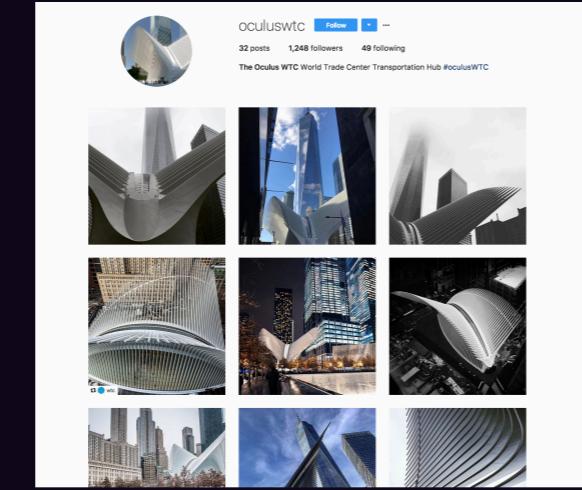
Who's on First is what's called a gazetteer.

That's map jargon for "a big list of places."

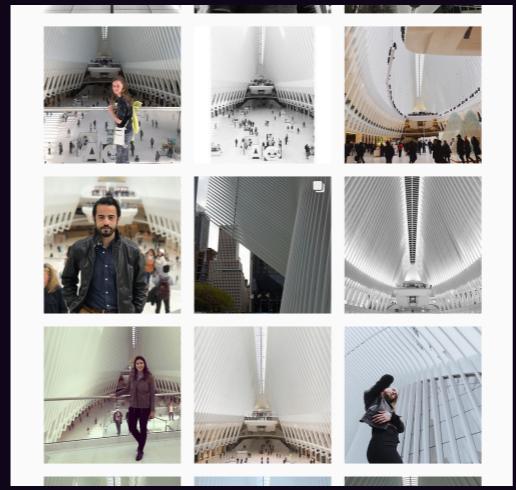
Some interesting properties we track for each place include:

- when it came into being, when it ceased to exist
- the hierarchies of places it sits inside of
- if there is a more relevant record that supersedes it

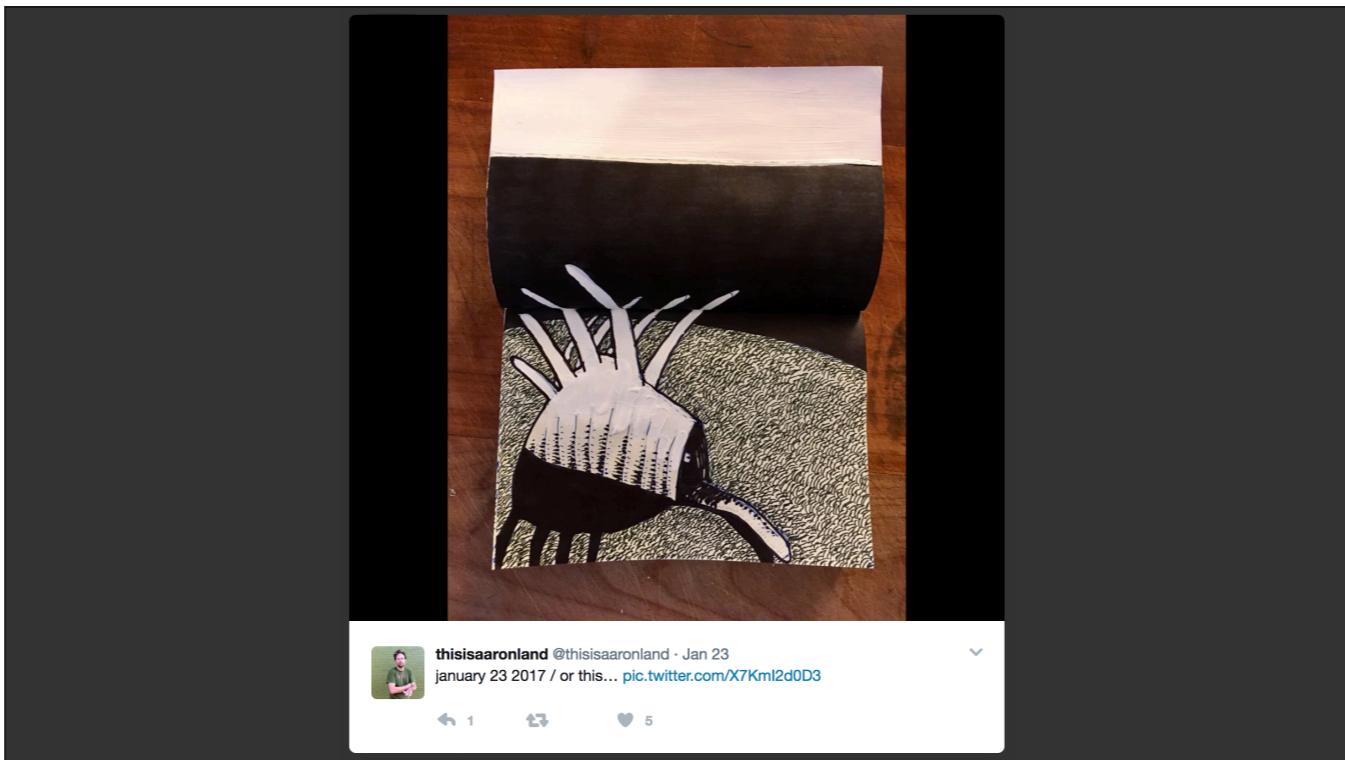
@oculuswtc
mostly external views



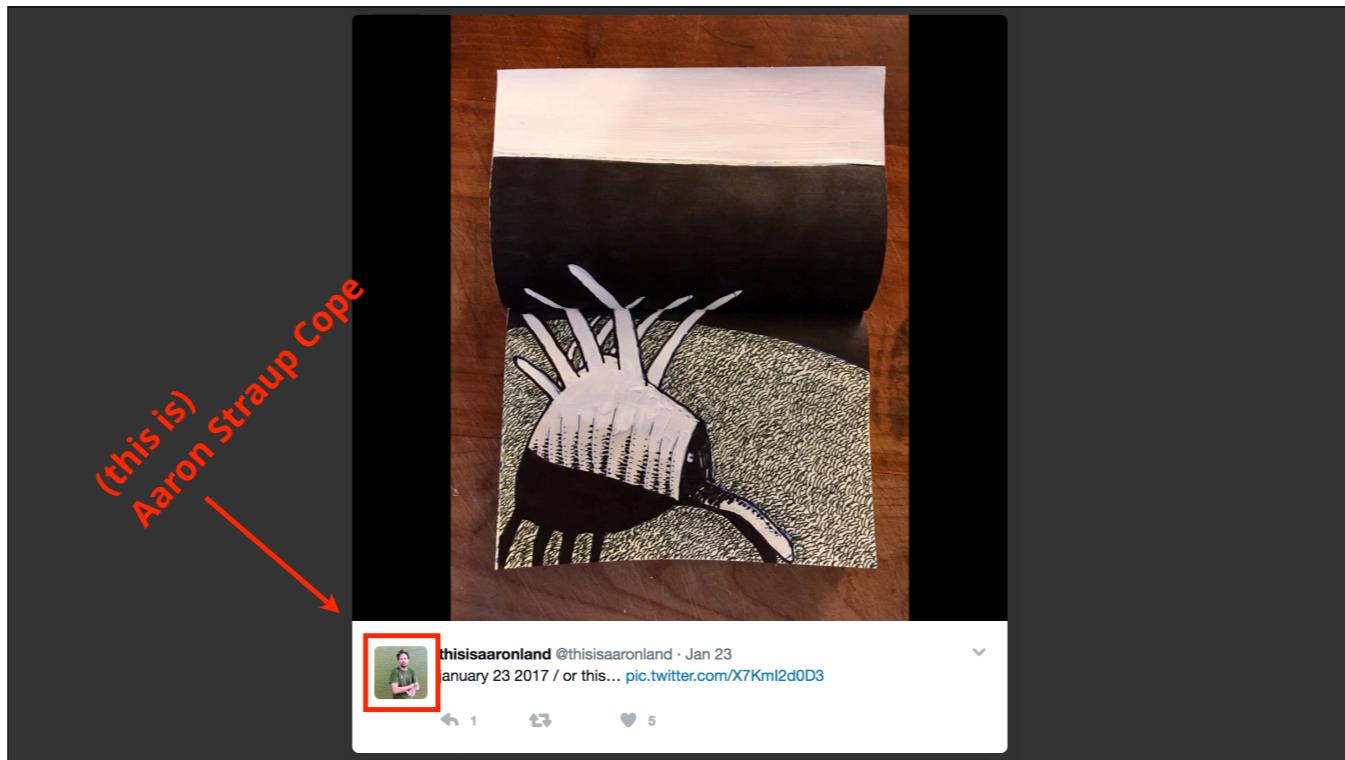
📍 wtc oculus
mostly internal views



Whether we see this place through the eyes of the official Instagram account or through the emergent place-tagged version, we can all agree that **this building exists**.



Who's On First represents human perspectives.
It also reflects our biases as the creators of the dataset.



Most of WOF's design originates from Aaron's long engagement working with open geo data.

```
{  
  id: 1108960835, ←  
  type: "Feature",  
  properties: {  
    "edtf:cessation": "uuuu",  
    "edtf:inception": "uuuu",  
    "geom:area": 0,  
    "geom:bbox": "-74.011406,40.71166,-74.011406,40.71166",  
    "geom:latitude": 40.71166,  
    "geom:longitude": -74.011406,  
    "iso:country": "",  
    "mz:categories": [ ],  
    "mz:hierarchy_label": 1,  
    "mz:hours": [ ],  
    "mz:is_current": 1,  
    "name:eng_x_colloquial": [  
      "The Carcass"  
    ],  
    "name:eng_x_preferred": [  
      "World Trade Center Transportation Hub"  
    ],  
    "name:eng_x_variant": [  
      "The Oculus"  
    ],  
    "src:geom": "mapzen",  
    "wof:belongsto": [  
      85865711,  
      102191575,  
      907196817,  
      85633793,  
      85977539,  
      421205771  
    ]  
  }  
}
```

Since each place has a stable WOF ID, we can start to build a network of place relationships.

```
{  
  id: 1108960835,  
  type: "Feature",  
  properties: {  
    "edtf:cessation": "uuuu",  
    "edtf:inception": "uuuu",  
    "geom:area": 0,  
    "geom:bbox": "-74.011406,40.71166,-74.011406,40.71166",  
    "geom:latitude": 40.71166,  
    "geom:longitude": -74.011406,  
    "iso:country": "",  
    "mz:categories": [ ],  
    "mz:hierarchy_label": 1,  
    "mz:hours": [ ],  
    "mz:is_current": 1,  
    "name:eng_x_colloquial": [  
      "The Carcass"  
    ],  
    "name:eng_x_preferred": [  
      "World Trade Center Transportation Hub"  
    ],  
    "name:eng_x_variant": [  
      "The Oculus"  
    ],  
    "src:geom": "mapzen",  
    "wof:belongsto": [  
      85865711,  
      102191575,  
      907196817,  
      85633793,  
      85977539,  
      421205771  
    ]  
  }  
}
```

Since each place has a stable WOF ID, we can start to build a network of place relationships.



As you may know, the quality of open data can be mixed. (I mean, the same is true with proprietary data too).
Our motto is **better than yesterday**.

who's on first jump to a random place

eliot center Search ▾

remember — this is a simple full-text search tool and not an [actual geocoder](#)

There are 17 places matching “eliot center”

```
{"query": {"function_score": {"boost_mode": "multiply", "functions": [{"filter": {"term": {"names_preferred": "eliot center"}, "weight": 3.0}, "filter": {"term": {"names_all": "eliot center"}, "weight": 1.0}, {"filter": {"term": {"wof:name": "eliot center"}, "weight": 1.5}, "filter": {"not": {"term": {"wof:placetype": "venue"}}, "weight": 2.0}, {"filter": {"exists": {"field": "wk:population"}}, "weight": 1.25}}]}}
```

Would you like try this query again to include records that have been deprecated ?



When I started prepping my talk I noticed that Who's On First doesn't know of an Eliot Center in Portland Oregon.

who's on first jump to a random place

eliot center Search ▾

remember — this is a simple full-text search tool and not an [actual geocoder](#)

There are 17 places matching “eliot center”

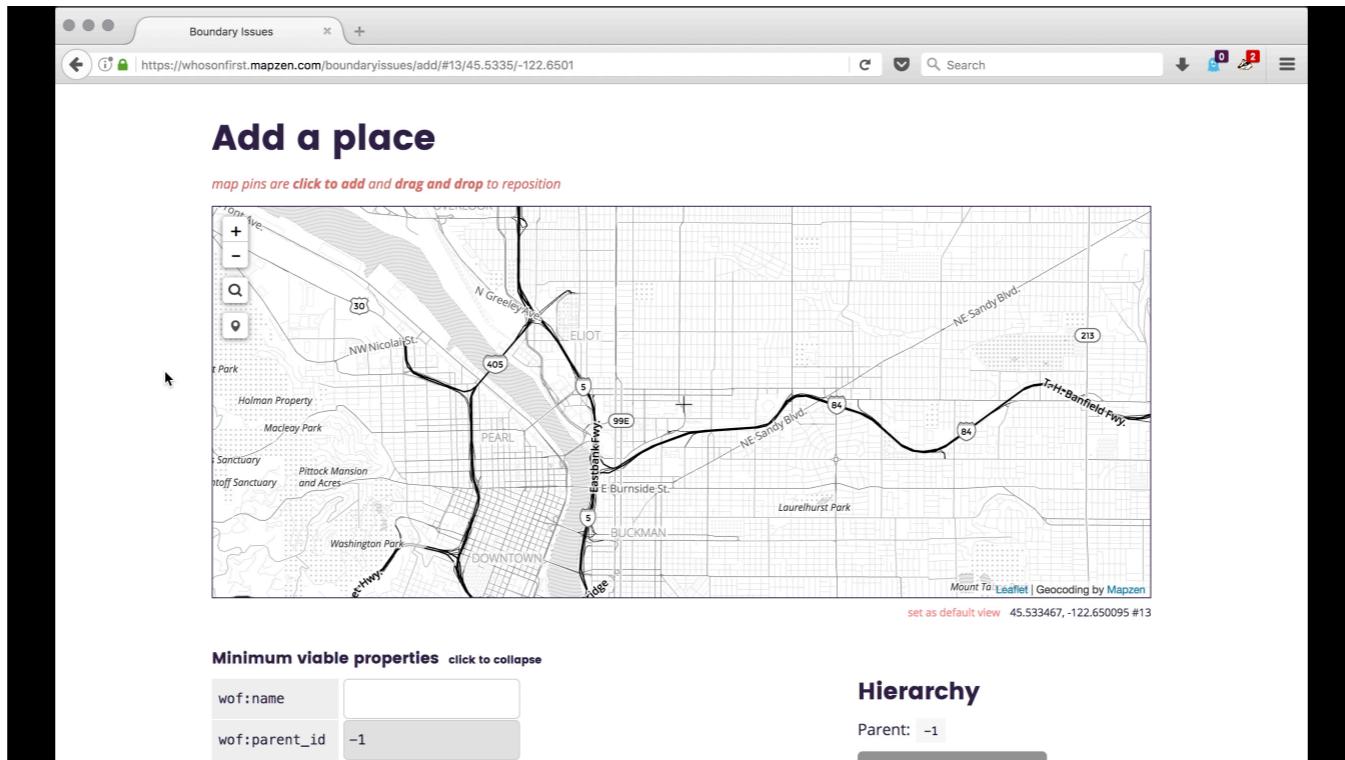
```
{'query': {'function_score': {'boost_mode': 'multiply', 'functions': [{ 'filter': { 'term': { 'names_preferred': u'eliot center' }}, 'weight': 3.0}, { 'filter': { 'term': { 'names_all': u'eliot center' }}, 'weight': 1.0}, { 'filter': { 'term': { 'wof:name': u'eliot center' }}, 'weight': 1.5}, { 'filter': { 'not': { 'term': { 'wof:placetype': 'venue' } }}, 'weight': 2.0}, { 'filter': { 'exists': { 'field': 'wk:population' }}, 'weight': 1.25}]}}, 'sort': { 'score': { 'order': 'desc' }}}
```

Would you like try this query again to include records that have been deprecated ?

+ -

Vancouver
Seattle
OR
ID
WY
SD
ND
NE
IA
MO
Minneapolis
WI
MI
Chicago
Algonquin Provincial Park
Toronto
Detroit
Boston
New York
St. Louis
IL
IN
OH
PA
WV
VA
NC
SC
MS
LA
TX
OK
AR
Dallas
Phoenix
San Francisco
Los Angeles
United States of America
Bermuda
SARGASSO SEA
Tampa
Houston

When I started prepping my talk I noticed that Who's On First doesn't know of an Eliot Center in Portland Oregon.



This is Boundary Issues, a web-based editor that I'm developing for Who's On First.
It is our bespoke CMS for editing WOF records.

I just want to pause and acknowledge there are a lot of moving parts here. They make sense to use internally, but I would imagine they look overwhelming to someone new to what we're doing.

The screenshot shows a web application interface for managing boundary issues. On the left, under 'Minimum viable properties', there are three fields: 'wof:name' (empty), 'wof:parent_id' (set to 85867131), and 'wof:placetype' (set to 'venue'). Below this is a section for 'Add a new property' with a 'Key' and 'Value' input field. Under 'Wof' (click to collapse), there are several fields: 'wof:concordances' (empty), 'wof:hierarchy' (set to {"continent_id":1021}), 'wof:name' (empty), 'wof:parent_id' (set to 85867131), 'wof:placetype' (set to 'venue'), 'wof:superseded_by' (button to 'Add an item'), 'wof:supersedes' (button to 'Add an item'), and 'wof:supersedes' (button to 'Add an item'). On the right, the 'Hierarchy' section shows a tree structure starting from 'Downtown' (id 85867131) with its children: 'North America' (id 102191575), 'United States' (id 85633793), 'Multnomah County' (id 102081631), 'Portland' (id 101715829), 'Southwest' (id 1108714057), 'Downtown' (id 85867131), and 'Oregon' (id 85688513). There is a 'REBUILD HIERARCHY' button. Below it is a 'Category' section with a dropdown menu set to 'Choose'.

The upshot is that all we really need to add a new venue is a lat/lon coordinate pair and a name.

2016-08-04T21:15:22.610Z	ok	wof_save_pending_branch: master	12595	6d092040-44d9-0000-8918-5e44258e18c2
2016-08-04T21:15:22.463Z	ok	git_push	12595	46fa8562-1691-0000-8d35-1d5fdf9ff0ca
2016-08-04T21:14:46.286Z	ok	git_commit	12595	905a4171-bff0-0000-8c95-df8525cd14c0
2016-08-04T20:42:43.577Z	ok	git_add	12595	9cf2e48d-429b-0000-b83e-c3f20e77545c
2016-08-04T20:42:39.148Z	ok	wof_save_pending_apply_diff	12595	b4c86aa3-d053-0000-a60a-430e90e86900
2016-08-04T20:42:39.038Z	ok	wof_geojson_encode	12595	e3b67261-23e4-0000-b2c8-bb3154f998de
2016-08-04T20:42:38.950Z	ok	wof_save_pending_pull: master	12595	0d8d529d-3afc-0000-97e1-99013ba8d053
2016-08-04T20:42:38.945Z	fail	git_pull: master	12595	26755f30-7d5a-0000-bb22-96891156e97f
2016-08-04T20:42:38.929Z	ok	git pull --rebase origin master	12595	8f7aa061-67e8-0000-be3c-2ca6b35191b4
2016-08-04T19:28:09.024Z	ok	git_curr_branch	12595	820f0192-5c99-0000-9fa6-eb96ba5cd35a
2016-08-04T19:28:08.917Z	ok	git_branches	12595	931742fa-0014-0000-896a-f0f7f2ce2a44

And then, behind the scenes: **stuff happens.**

The screenshot shows a GitHub commit page for the repository `whosonfirst-data / whosonfirst-data-venue-us-or`. The commit was made by `dphiffer` on the `master` branch. The commit message is `* Eliot Center (1108955793) saved by dphiffer`. The commit was pushed 7 minutes ago and has 1 parent, with a commit hash of `5bad89b70d34ac26c6418d28d2d3080cc98e4011`. The file changed is `data/110/895/579/3/1108955793.geojson`, showing 66 additions and 0 deletions. The code added is:

```
66 1108955793.geojson
@@ -0,0 +1,66 @@
1 +{
2 + "id": 1108955793,
3 + "type": "Feature",
4 + "properties": {
5 +   "addr:full": "1034 Sw 13th Ave, Portland, OR, USA",
6 +   "addr:housenumber": "1034",
7 +   "addr:postcode": "97205",
8 +   "addr:street": "Sw 13th Ave",
9 +   "edtf:cessation": "uuuu",
10 +  "edtf:inception": "uuuu",
11 +  "geom:area": 0.0,
12 +  "geom:bbox": "-122.686174,45.51821,-122.686174,45.51821",
13 +  "geom:latitude": 45.51821,
14 +  "geom:longitude": -122.686174,
```

And now this place exists in the data set.

Boundary Issues translates your web-based edits into git-speak.

Who's On First

- Open Data gazetteer
- Omnivorous sourcing (all open-licensed)
- Stable identifiers (`wof:id` property)
- Concordances to other dataset IDs
- Encoded as GeoJSON flat files
- Everything flows through GitHub
- Extensible “semantic web lite” property encoding

Here is a recap of what Who's On First is about.

Who's On First

- Open Data gazetteer
- Omnivorous sourcing (all open-licensed)
- Stable identifiers (`wof:id` property)
- Concordances to other datasets's IDs
- Encoded as GeoJSON flat files
- Everything flows through GitHub
- Extensible “semantic web lite” property encoding

I'm going to be focusing on that last bullet point for the rest of this talk.

And I'm thinking about how user interface influences who's able to participate in defining places.

Boundary Issues

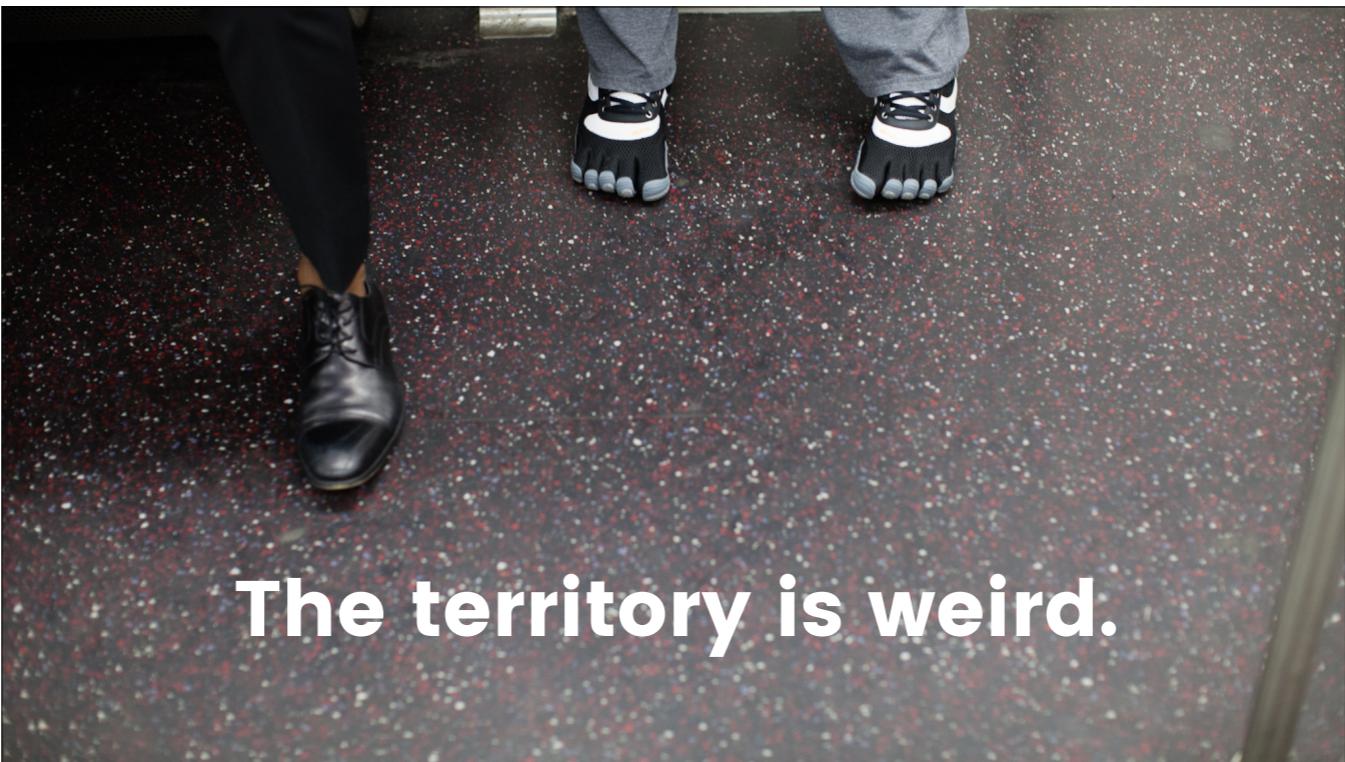
- Our CMS for editing Who's On First
- Translates web-based edits into git commits
- Works in tandem with other tools like QGIS
- Still not available to the public
- Just started inviting outside people to test things out
- Expands who's able to participate

And this is Boundary Issues.

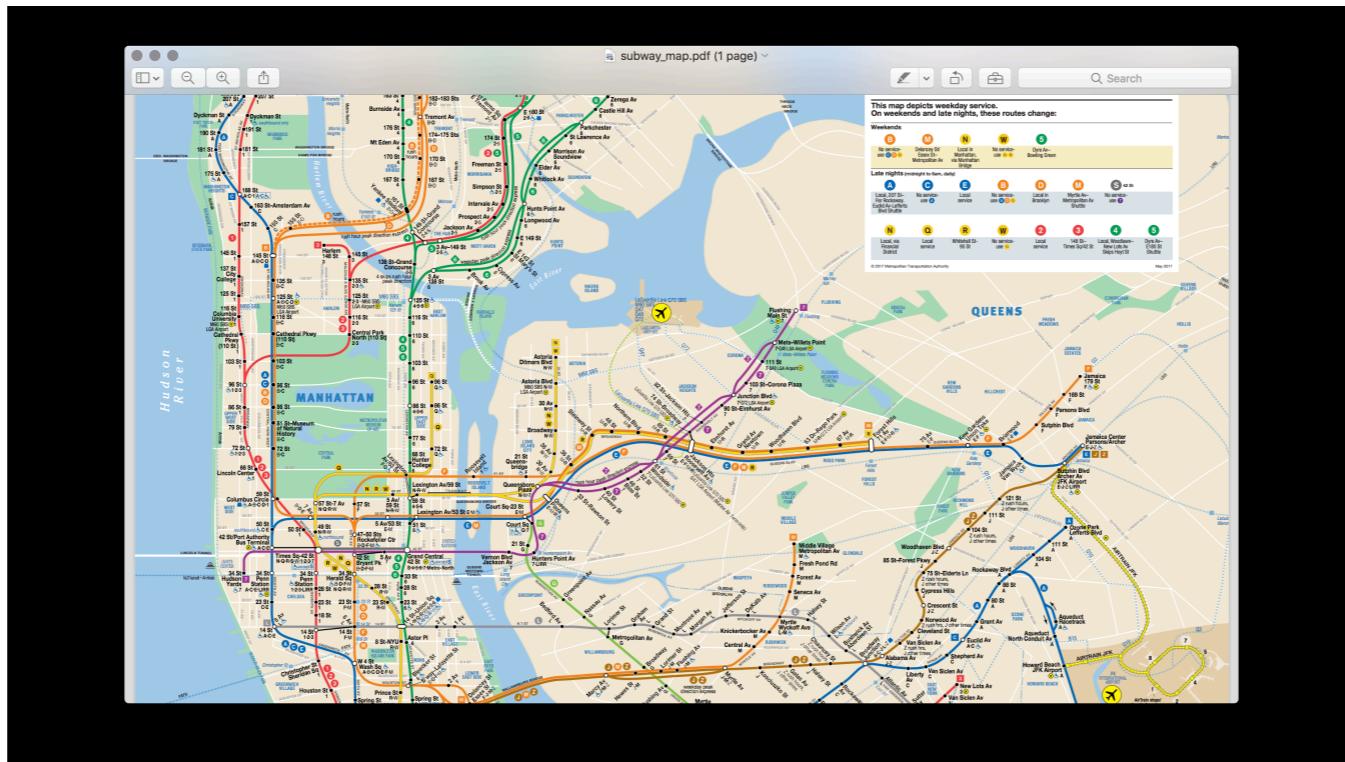
Boundary Issues

- Our CMS for editing Who's On First
- Translates web-based edits into git commits
- Works in tandem with other tools like QGIS
- Still not available to the public
- Just started inviting outside people to test things out
- Expands who's able to participate

I'm thinking a lot about how user interface influences who's able to participate in defining places.



It's easy enjoy weird maps because the territory is weird.



Like many New Yorkers who grew up on the west coast, the subway system is endlessly fascinating to me.



Our sprawling 24-hour transit is a source of pride, the subject of gripes and strongly held opinions.



And maybe the most interesting venue for contemporary dance in a city well known for its performing arts.

Wheelchair accessible subway map

"While there are **490 stations** in the New York system (including Staten Island), barely more than **100 stations are accessible**"

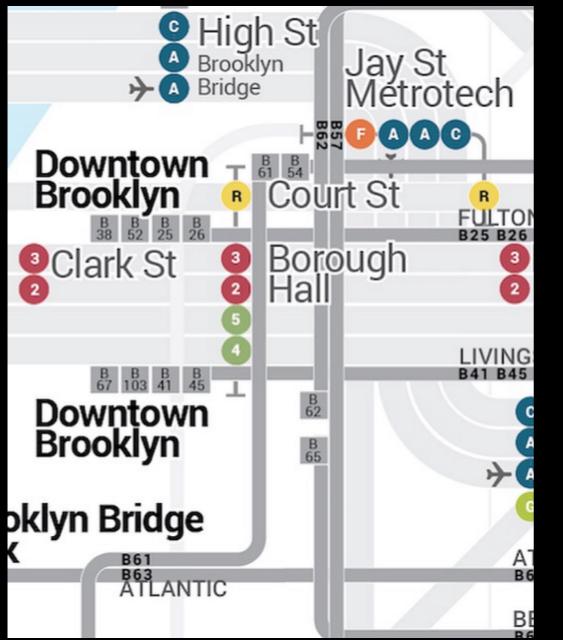
—Matthew Ahn



The “subway map” means different things to different people. And that is because the subway itself means different things to different people.

Bullet Map by Anthony Denaro

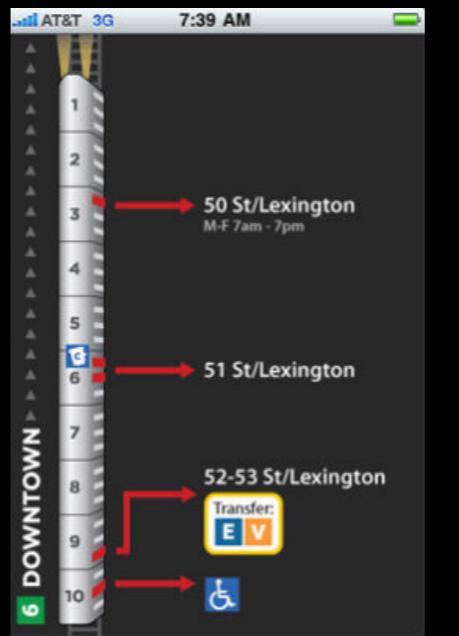
"Why aren't local buses, crosstown buses and the new fancy SBS buses shown? Could **the whole bus system** fit over a subway map, especially in the areas that aren't served by the subway?"



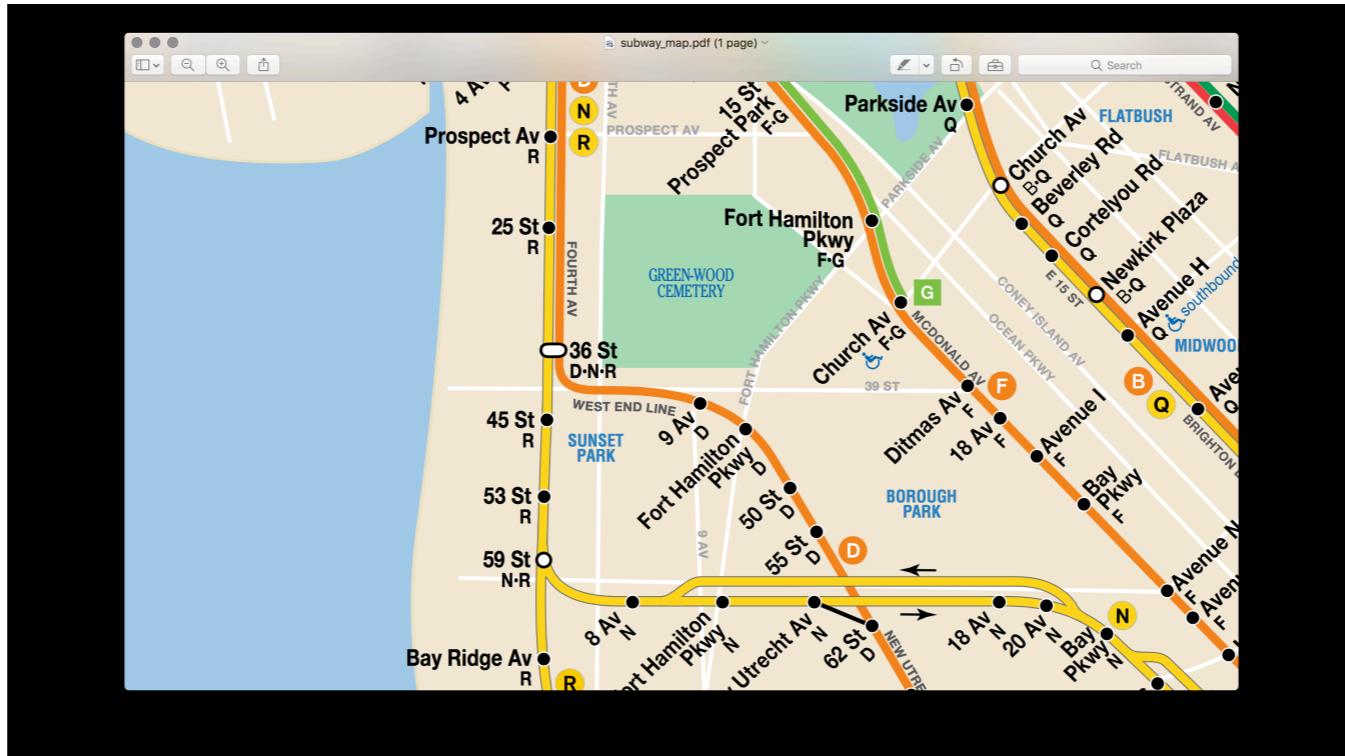
The “subway map” means different things to different people. And that is because the subway itself means different things to different people.

NYC Exit Strategy

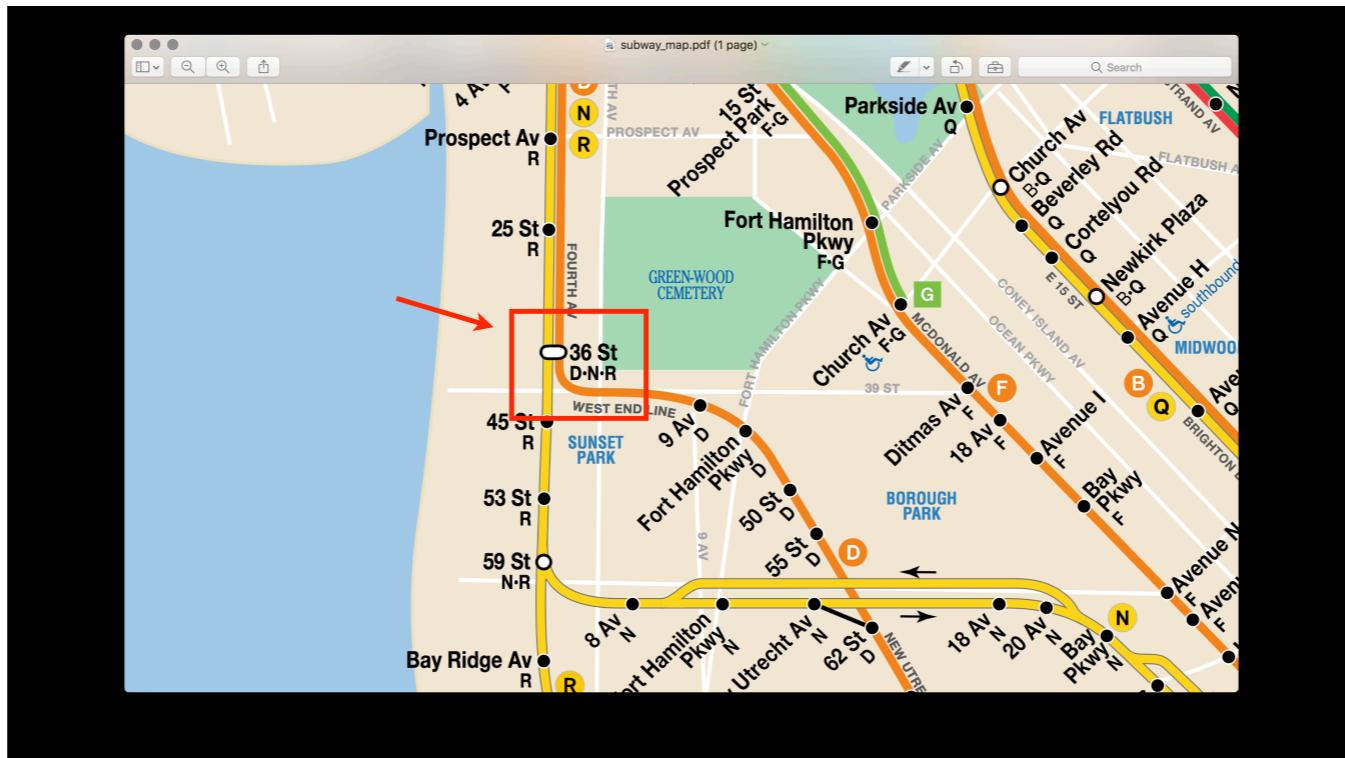
"Tells you precisely which subway car to ride in so you get off **exactly** where you want to, as close to the right exit as possible."



The “subway map” means different things to different people. And that is because the subway itself means different things to different people.



Let's isolate ourselves to the 36th St station in the historically working class neighborhood of Sunset Park.

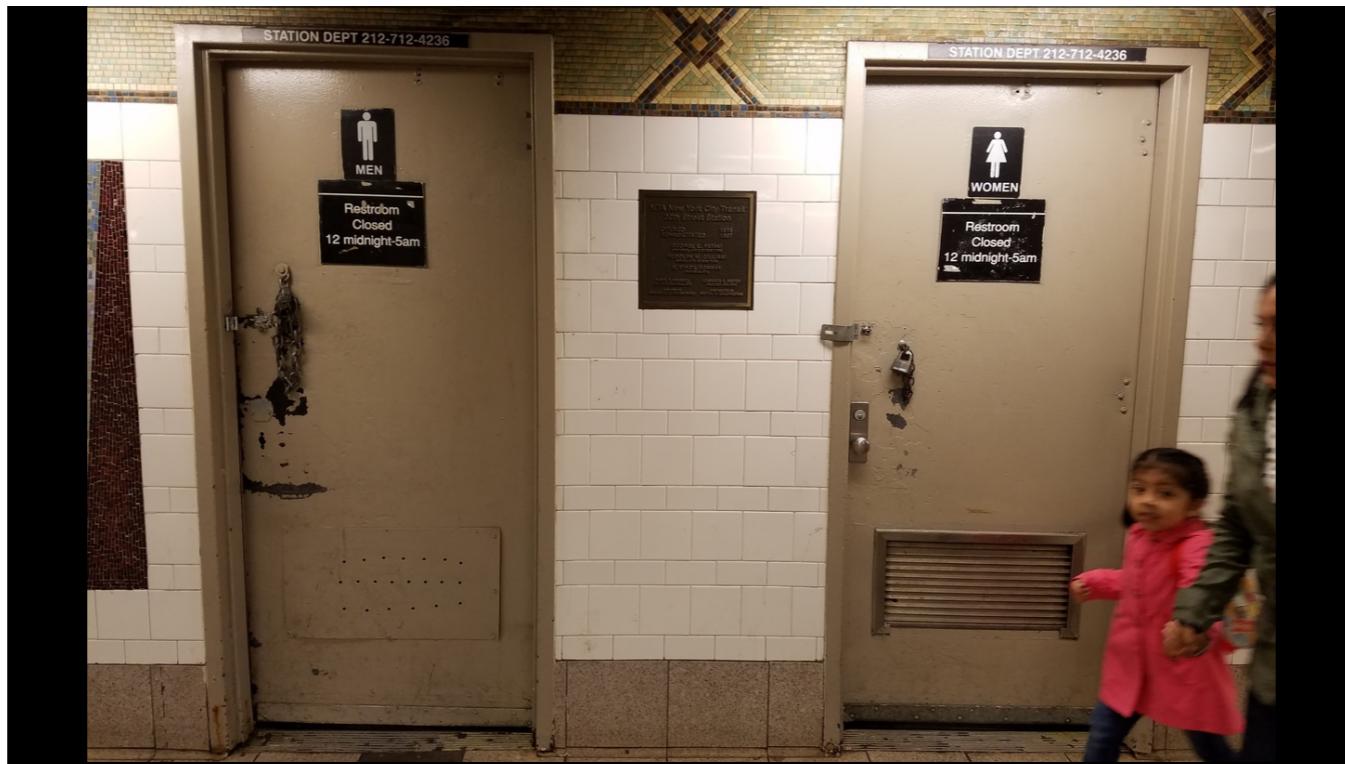


Let's isolate ourselves to the 36th St station in the historically working class neighborhood of Sunset Park.



You can find this useful metadata embedded within the station itself.

Let's just stand back a little bit.



Behold, there are restrooms here! (For those not familiar with NYC subway facilities, this is uncommon.)
Let's keep looking around this station.



Check out these tile murals!



They depict the designers, workers...



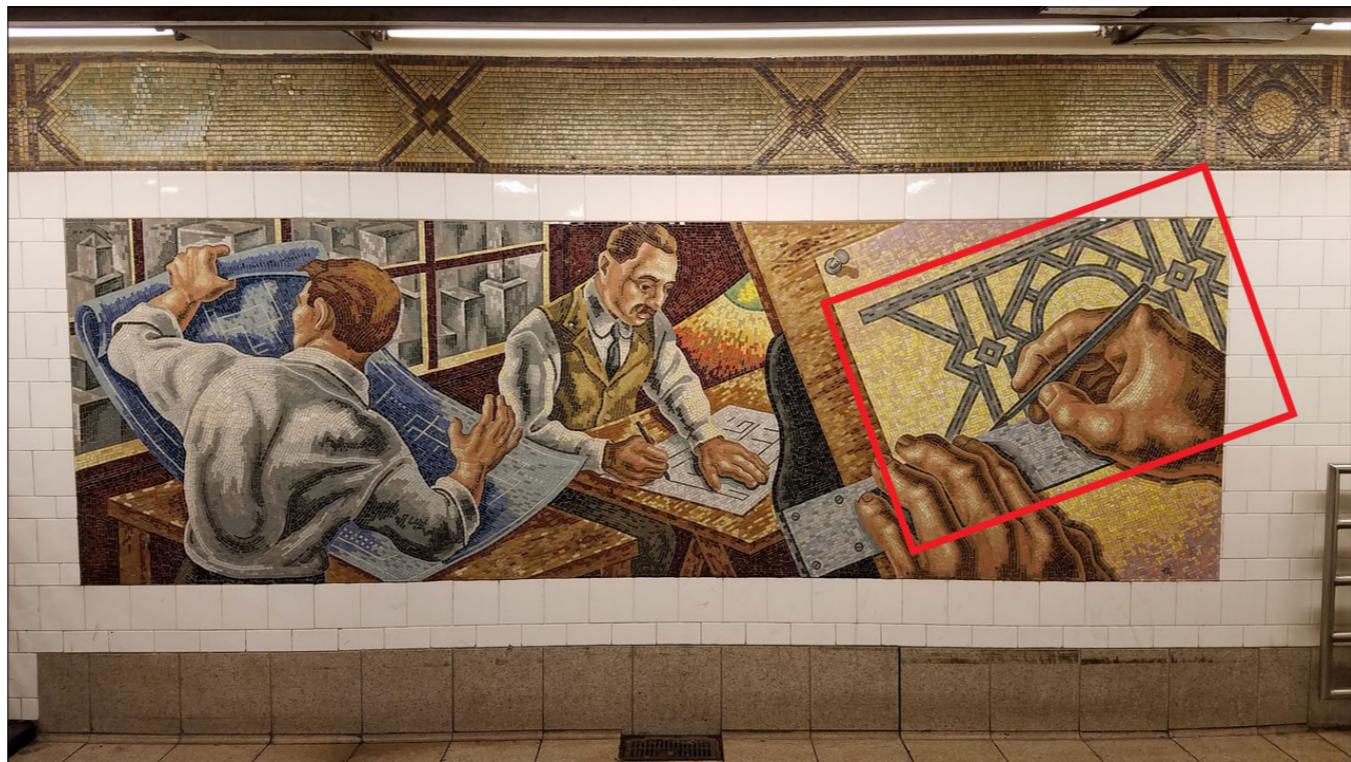
... and riders.



Like all of the public art in NYC subway stations, we've got more good metadata here.



Did you catch this part?



Did you catch this part?



The mural commissioned in 1998 references tile work that may date back to when the station opened in 1915.

Station layout [edit]			Track layout [hide]	Address
G	Street Level	Exit/Entrance		36th Street & Fourth Avenue Brooklyn, NY 11232
M	Mezzanine	Fare control, station agent		Borough Brooklyn
P		← D toward Norwood–205th Street, N toward Ditmars Boulevard (late nights) (25th Street)	<p>Legend</p> <p>to Atlantic Av to 25 St</p>	Locale Sunset Park
Platform level	Northbound local	← R toward Forest Hills–71st Avenue (Whitehall Street) late nights (25th Street) ← W toward Ditmars Boulevard rush hours (25th Street)		Coordinates 40.65445°N 74.00455°W
		Island platform, doors will open on the left, right		Division B (BMT)
		← D toward Norwood–205th Street (Atlantic Avenue–Forest Hills–71st Avenue–Barclays Center)		Line BMT Fourth Avenue Line
		→ D toward Coney Island–Stillwell Avenue via West End (Ninth Avenue) → express		Services
		N toward Coney Island–Stillwell Avenue via Sea Beach (59th Street) →	<ul style="list-style-type: none"> D ● (all times) N ● (all times) R ● (all times) W ♦ (limited rush hour service only) 	
		Island platform, doors will open on the left, right		Transit connections
		D N toward Coney Island–Stillwell Avenue (late nights) (Ninth Avenue (D) or 45th Street (N)) → R toward Bay Ridge–95th Street (45th Street) → W toward 86th Street–Gravesend rush hours (45th Street) →		NYCT Bus: B70
				Platforms
				cross-platform interchange
				Tracks
				4
				Other information
				Opened June 22, 1915 (101 years ago)
				Wireless service
				Traffic
				Passengers 4,138,929 [2] 4.2% (2015)
				Rank 121 out of 425
				Station succession
				Next north 25th Street (local): D N R W ♦

What counts as “data?”

Wikipedia has plenty of data about 36th St, but does not mention restrooms nor public art.

Lots of the nuance that captures why we should even care about places tends to get ignored by datasets that are considered comprehensive.

```
{  
  id: 571779669,  
  type: "Feature",  
  properties: {  
    "addr:full": "Broadway & Murray St New York NY 10007",  
    "addr:housenumber": "",  
    "addr:postcode": "10007",  
    "addr:street": "St",  
    "edtf:cessation": "uuuu",  
    "edtf:inception": "uuuu",  
    "geom:area": 0.0,  
    "geom:area_square_m": 0.0,  
    "geom:bbox": "-74.0072107315,40.7130450143,-74.0072107315,40.7130450143",  
    "geom:latitude": 40.713045,  
    "geom:longitude": -74.007211,  
    "iso:country": "US",  
    "mz:hierarchy_label": 1,  
    "sg:address": "Broadway & Murray St",  
    "sg:city": "New York",  
    "sg:classifiers": [  
      {  
        category: "Local Transit",  
        subcategory: "",  
        type: "Transportation"  
      }  
    ],  
    "sg:owner": "simplegeo",  
    "sg:postcode": "10007",  
    "sg:province": "NY",  
    "sg:tags": [  
      "public"  
    ],  
    "src:geom": "transitland",  
    "wof:belongsto": [  
  ]
```

Here you see a record for a subway stop. Notice how each property key has a prefix? It's not just "tags," it's "**sg** tags." The "sg" prefix is an indication that this property originated from the SimpleGeo dataset.

```
{  
  id: 571779669,  
  type: "Feature",  
  properties: {  
    "addr:full": "Broadway & Murray St New York NY 10007",  
    "addr:housenumber": "",  
    "addr:postcode": "10007",  
    "addr:street": "St",  
    "edtf:cessation": "uuuu",  
    "edtf:inception": "uuuu",  
    "geom:area": 0.0,  
    "geom:area_square_m": 0.0,  
    "geom:bbox": "-74.0072107315,40.7130450143,-74.0072107315,40.7130450143",  
    "geom:latitude": 40.713045,  
    "geom:longitude": -74.007211,  
    "iso:country": "US",  
    "mz:hierarchy_label": 1,  
    "sg:address": "Broadway & Murray St",  
    "sg:city": "New York",  
    "sg:classifiers": [  
      {  
        category: "Local Transit",  
        subcategory: "",  
        type: "Transportation"  
      }  
    ],  
    "sg:owner": "simplegeo",  
    "sg:postcode": "10007",  
    "sg:province": "NY",  
    "sg:tags": [  
      "public"  
    ],  
    "sg:geom": "transitland",  
    "wof:belongsto": [  
    ]  
  }  
}
```

This is a WOF record for a subway stop. Notice how each property key has a prefix? It's not just "tags," it's **sg** tags." The "sg" prefix is an indication that this property originated from the SimpleGeo dataset.

simplegeo.json

<https://github.com/whosonfirst/whosonfirst-sources/blob/master/sources/simplegeo.json>

```
{  
  id: 404734199,  
  fullname: "SimpleGeo",  
  name: "simplegeo",  
  prefix: "sg",  
  key: "id",  
  url: "",  
  license: https://creativecommons.org/publicdomain/zero/1.0/,  
  description: "SimpleGeo was a location aware services company that  
  operated between 2009 and 2011. It is no longer an active company."  
}
```

Property prefixes gives us a useful way to group those needs, avoid naming conflicts, and let them all cohabitate in a single WOF record.

libraryofcongress.json

<https://github.com/whosonfirst/whosonfirst-sources/blob/master/sources/libraryofcongress.json>

```
{  
  id: 840464301,  
  fullname: "Library of Congress",  
  name: "loc",  
  prefix: "loc",  
  key: "id",  
  url: http://www.loc.gov,  
  license: https://www.usa.gov/government-works  
}
```

Some sources are authoritative.

burritojustice.json

<https://github.com/whosonfirst/whosonfirst-sources/blob/master/sources/burritojustice.json>

```
{  
  id: 404734205,  
  fullname: "Burrito Justice",  
  name: "burritojustice",  
  prefix: "bj",  
  key: "id",  
  url: http://burritojustice.com/la-lengua/,  
  license: "CC0",  
  description: ""  
}
```

Some are more vernacular. John Oram aka burritojustice has helped define some key microhoods of the Mission in SF.



John also made this map, which was recently featured in the New York Times Magazine.

The world is wonderfully multifaceted.

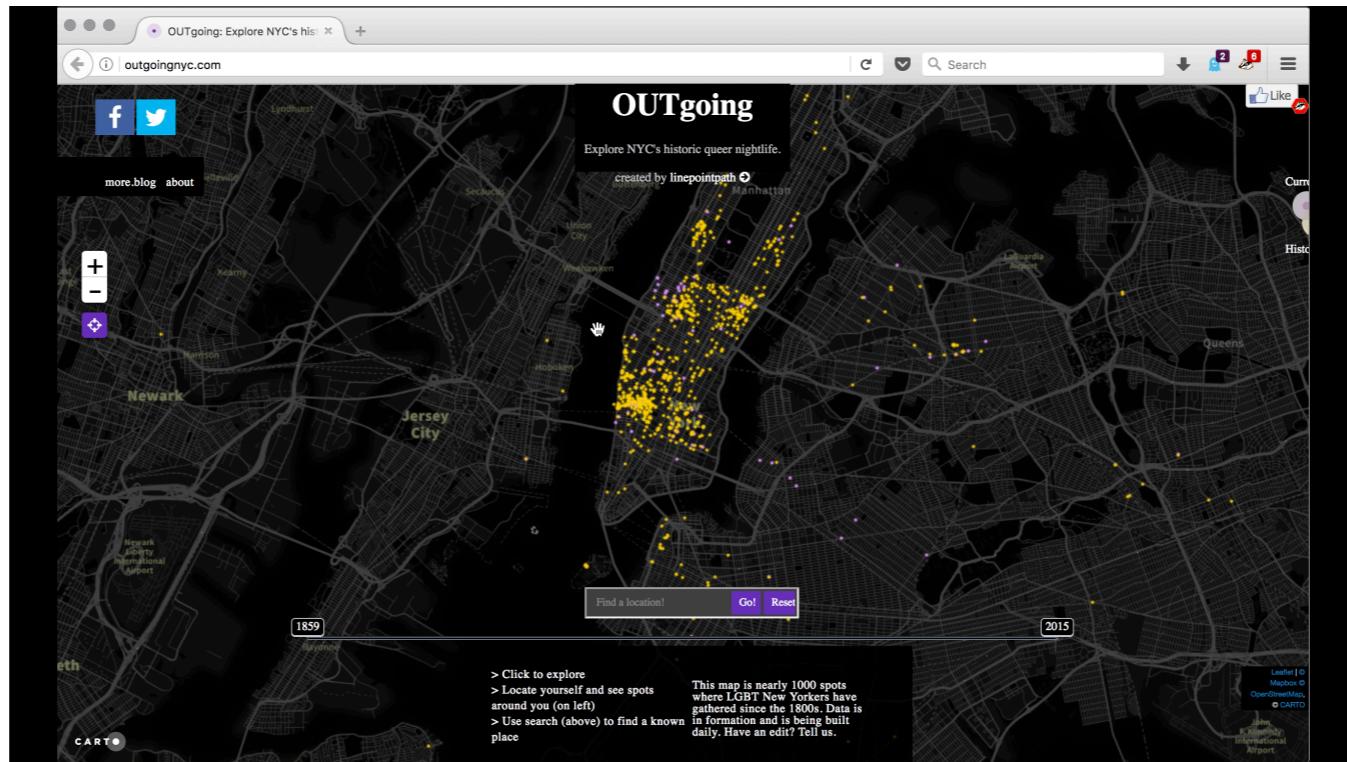
<https://twitter.com/Walldo/status/852040146446831616>

The territory means different things to different people. The territory also means different things to different species.

Depending on your perspective, the kinds of data that are captured about places may be missing, insufficient, or downright hostile.

Who's On First is **opinionated**—like all datasets, no collection is truly unbiased—but we hope to be aware of when we're asserting our own opinions about places.

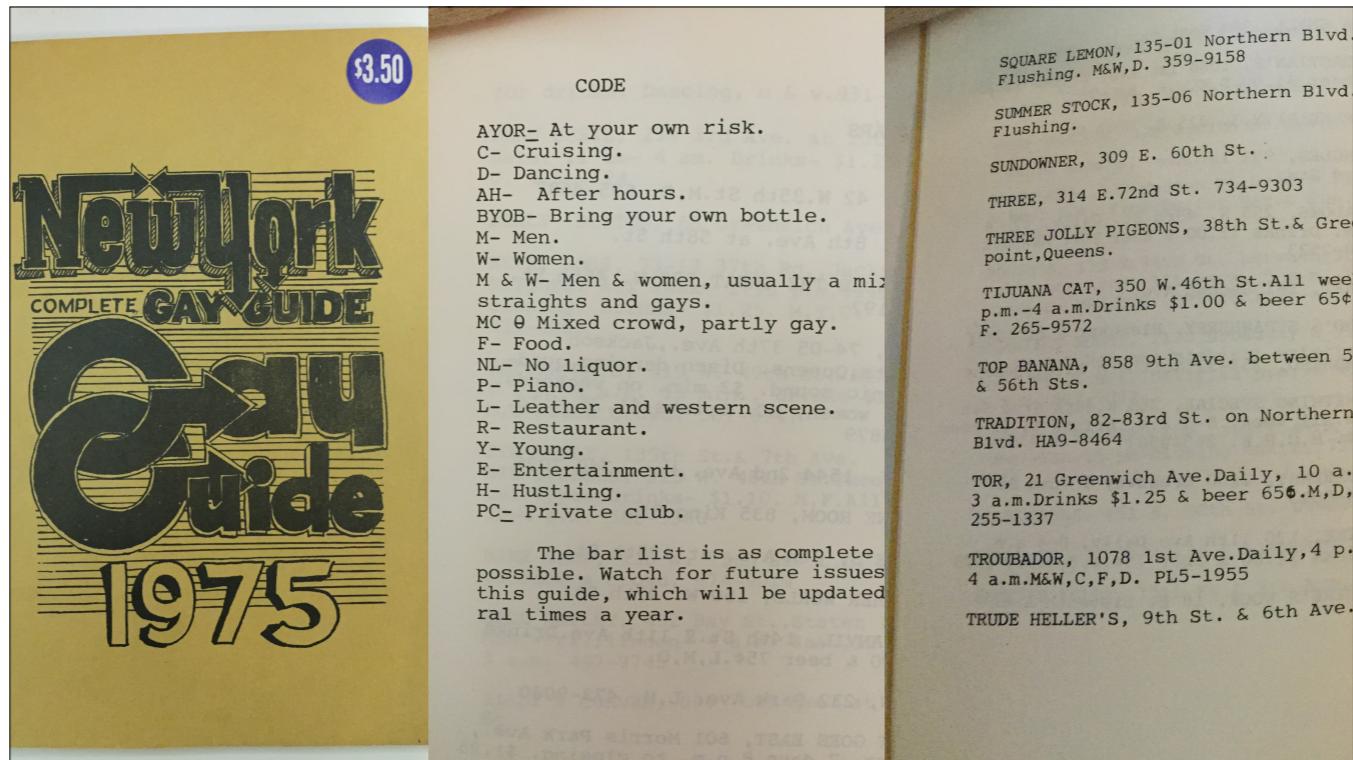
We are building a framework where a polyglot of **place-feels** will be welcome.



Like the Internet, Who's On First data is **made of people**.

Those who have the capability to edit our data necessarily reflect themselves in those edits.

They are present in the data, through choices about which places get edited, and the types of nuanced assertions they make about places.



During the development of our new CSV import feature, I had the pleasure of working with Jeff Ferzoco's dataset for his excellent historic mapping project, OUTgoing.

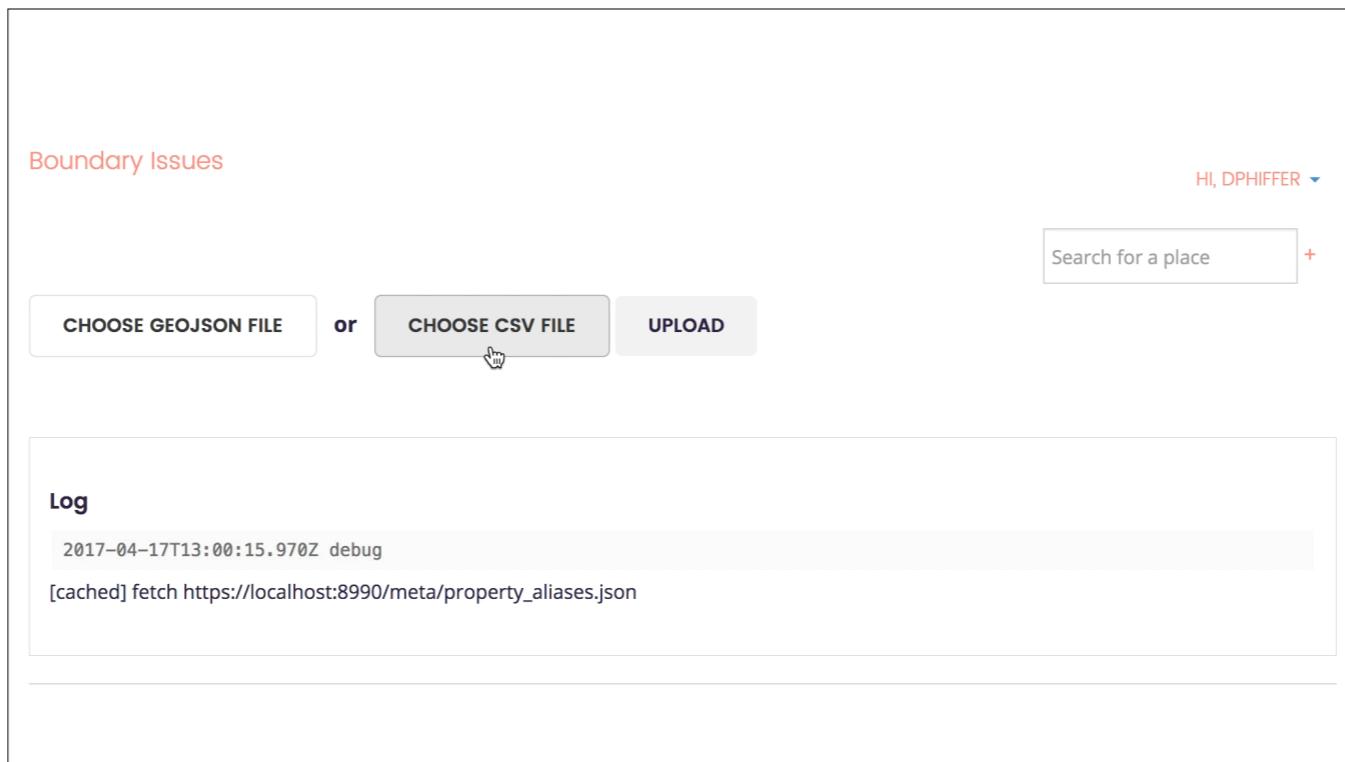
CSV outgoing.csv

Raw

Search this file...

in_business	genre	genre2	name	location	city
FALSE	Club	Dance	Paradise Garage	84 King Street	New York
FALSE	Club	Dance	Mais Oui		New York
FALSE	Club	Sex	Zone DK		New York
FALSE	Club	Dance	Androgyny	35 Crosby	New York
FALSE	Club	Sex	EI Mirage	253 Houston Street	New York
FALSE	Bar	Neighborhood	The Urge (v1)		New York
FALSE	Bar	Restaurant	Ninth Circle	139 West 10th	New York
FALSE	Bar	Neighborhood	Peppermint Lounge		New York
TRUE	Bar	Neighborhood	Julius'	159 W 10th St	New York
FALSE	Hotel		Van Rensselaer Hotel	15 East 11 St	New York
FALSE	Bar		Cherchez La Femme	108 West 43rd Street	New York
FALSE	Bathhouse		The Ariston	55th Street and Broadway	New York
FALSE	Bar	Bear	Dugout		New York
FALSE	Club	Sex	Show Palace	670 Eighth Av	New York

Jeff was kind enough to share a CSV file of all his collected places, which is dense with excellent metadata like **year_opened**, **year_closed**, and **mafia_owned**.



I used Jeff's dataset to test a new feature that lets us import data from CSV files. I'm gonna show you how that works.

First CSV row is column headers

Geometry source Known source: [Mapzen](#)

Property prefix

COLUMN HEADER	WOF PROPERTY	PREVIEW VALUES < 1 / 991 >
source	mz:source	http://www.nytimes.com/2000/06/18/style/paradise-garage-a-gay-club-that-forever-changed-night-life.html
in_business	mz:in_business	FALSE
genre	mz:genre	Club
genre2	mz:genre2	Dance

We can choose some general options about whether the CSV file has a **header row**, we can identify a **source** for the data (in this case, outgoing), and choose a **prefix** for the properties we are merging into Who's On First.

COLUMN HEADER	WOF PROPERTY	PREVIEW VALUES < 1 / 991 >
source	out:source	http://www.nytimes.com/2000/06/18/style/paradise-garage-a-gay-club-that-forever-changed-night-life.html
in_business	out:in_business	FALSE
genre	out:genre	Club
genre2	out:genre2	Dance
name	wof:name	Paradise Garage
location	out:location	84 King Street
city	addr:city	New York
state	addr:state	NY
full_location	out:full_location	84 King Street, New York, NY
latitude	geom:latitude	40.7279733
longitude	geom:longitude	-74.0064172

Next we will map **which columns** from the CSV should be saved to **which properties** in the WOF record.

Some columns map onto generic properties like **latitude**, and **longitude**, while others get saved as new properties with the **out:** prefix.

Import from CSV

◀ prev **1 of 991** next ▶ outgoing.csv

Venue name

Paradise Garage

Address

84 King Street, New York, NY

find it on the map

Tags (comma-separated)

Preview property assignments

ADD VENUE



Once that property mapping is ready, we can click the “upload” button and begin stepping through each of the rows in the CSV file.

Venue name

Zone DK

Address

[find it on the map](#)

Tags (comma-separated)

[Preview property assignments](#)



Import from CSV

◀ prev 554 of 991 next ▶ outgoing.csv

Does this record exist already? This seems similar to:
[The Stonewall Inn \(1 of 1\)](#)

SAME PLACE

NOT A DUPE

Venue name

Stonewall

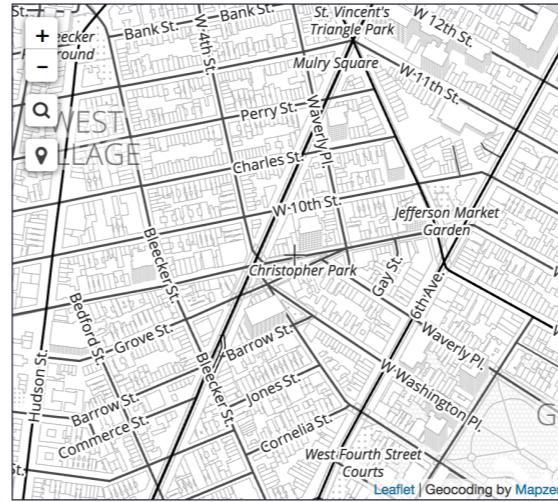
Address

53 Christopher St, New York, NY

find it on the map

Tags (comma-separated)

Bar, Gay



40.733794, -74.002155 #16

We also want to avoid duplicating existing records, if possible.

The screenshot shows a web application interface for managing boundary issues. At the top, there's a header with the text "Boundary Issues" and "HI, DPHIFFER". Below the header is a search bar with the placeholder "Search for a place" and a plus sign icon. The main content area has a title "Import from CSV" and a sub-instruction: "Download the your CSV file, modified with the imported WOF IDs (`wof_id`) inserted into the first column." There are two buttons: "MODIFIED CSV" (which is highlighted with a blue background) and "ORIGINAL CSV". Below these buttons is a "Log" section containing the following text:
2017-04-17T17:26:10.576Z debug
[cached] fetch https://localhost:8990/meta/property_aliases.json

Once we complete the step-through import process, we can download a new CSV file modified to include a new **wof_id** column. This additional column allows us to round-trip our edits back into Who's On First, effectively turning your spreadsheet editor into a new external interface to change properties en masse.

WOF ❤ CSV

<https://whosonfirst.mapzen.com/>
@mapzen

boundaryissues@mapzen.com
@alloftheplaces

dphiffer.org
@dphiffer

We look forward to meeting your weird and wonderful CSVs!