

If you can read this, you don't need glasses





What's a geocoder? It's a service that finds places, and lets things like this autocomplete box work their magic.

What does a geocoder have to do? It's really simple, only two things:



Know about every place in the world

 @juliansimioni 

Philadelphia

Philadelphia

Philadelphia

费城

ფილადელფია

פִּילָּדְּלָפִּיה

Philadelphia

Philadelphia

費城

費城

Fi-sàng

Philadelphia

Philadelphia

פִּילָּדְּלָפִּיה

फिलाडेल्फिया

Philadelphia

Philadelphia

Ֆիլադելֆիա

Philadelphia

Philadelphia

فیلadelفیا

Philadelphia

ഫിലാദേല്പിയ

ফিলাডেল্ফিয়া

Филадельфий

Филаделфија

Philadelphia

Филадельфия

Philadelphia

Филадельфи

Give people what they mean
no matter what they type

 @juliansimioni 

eng_x_colloquial

Phillie
Philly
Philthadelphia
Philthy
Quaker City
The City of Brotherly Love
City of Brotherly Love
Hostile City
Cradle of Liberty
The Quaker City
The Illadelph
The 215
PHL
Shackamaxon
The Big Scrapple
Filphadelphia
Filthydelphia
Stinkydelphia

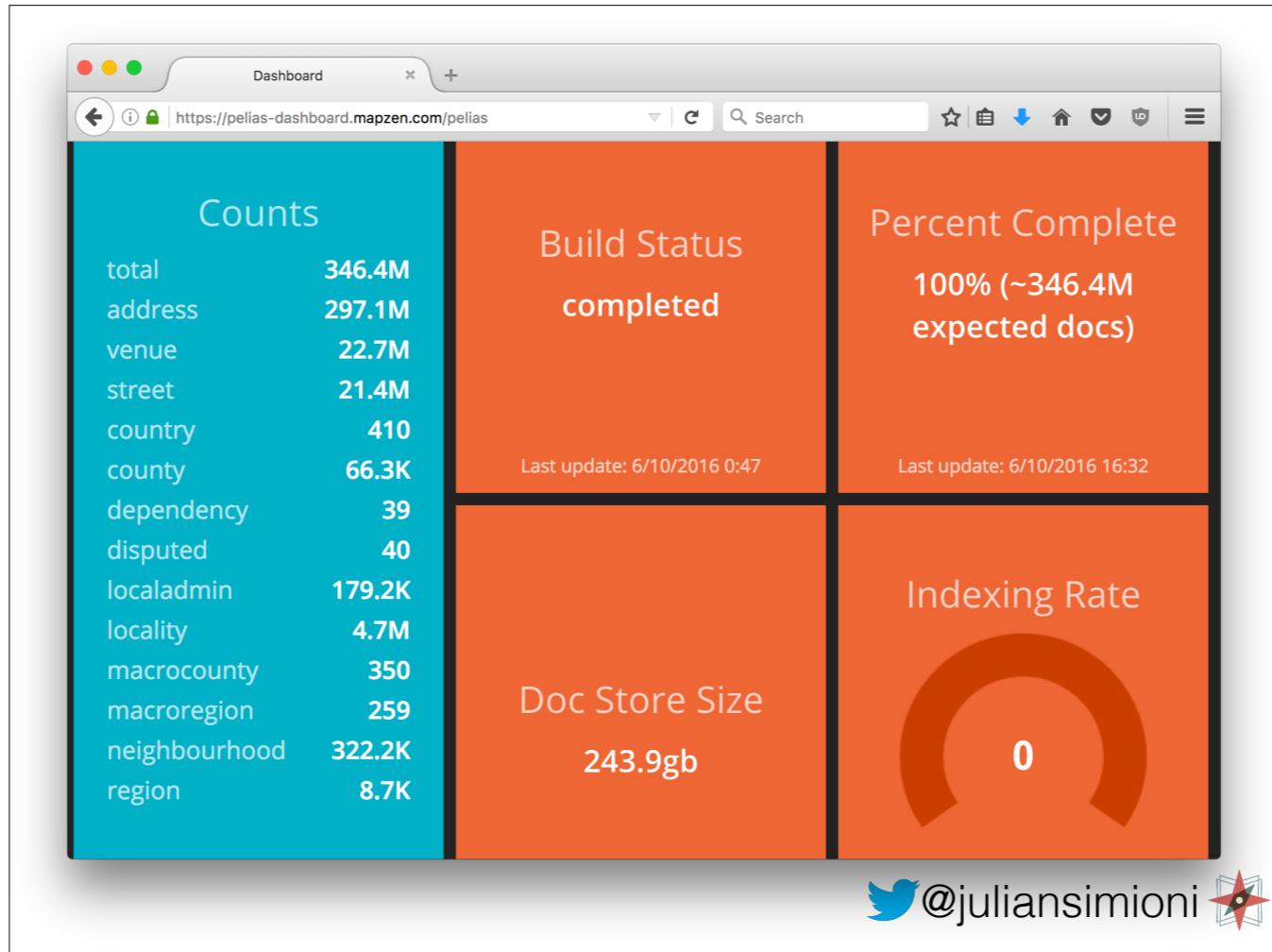
<https://whosonfirst.mapzen.com/spelunker/id/101718083>

 @juliansimioni 

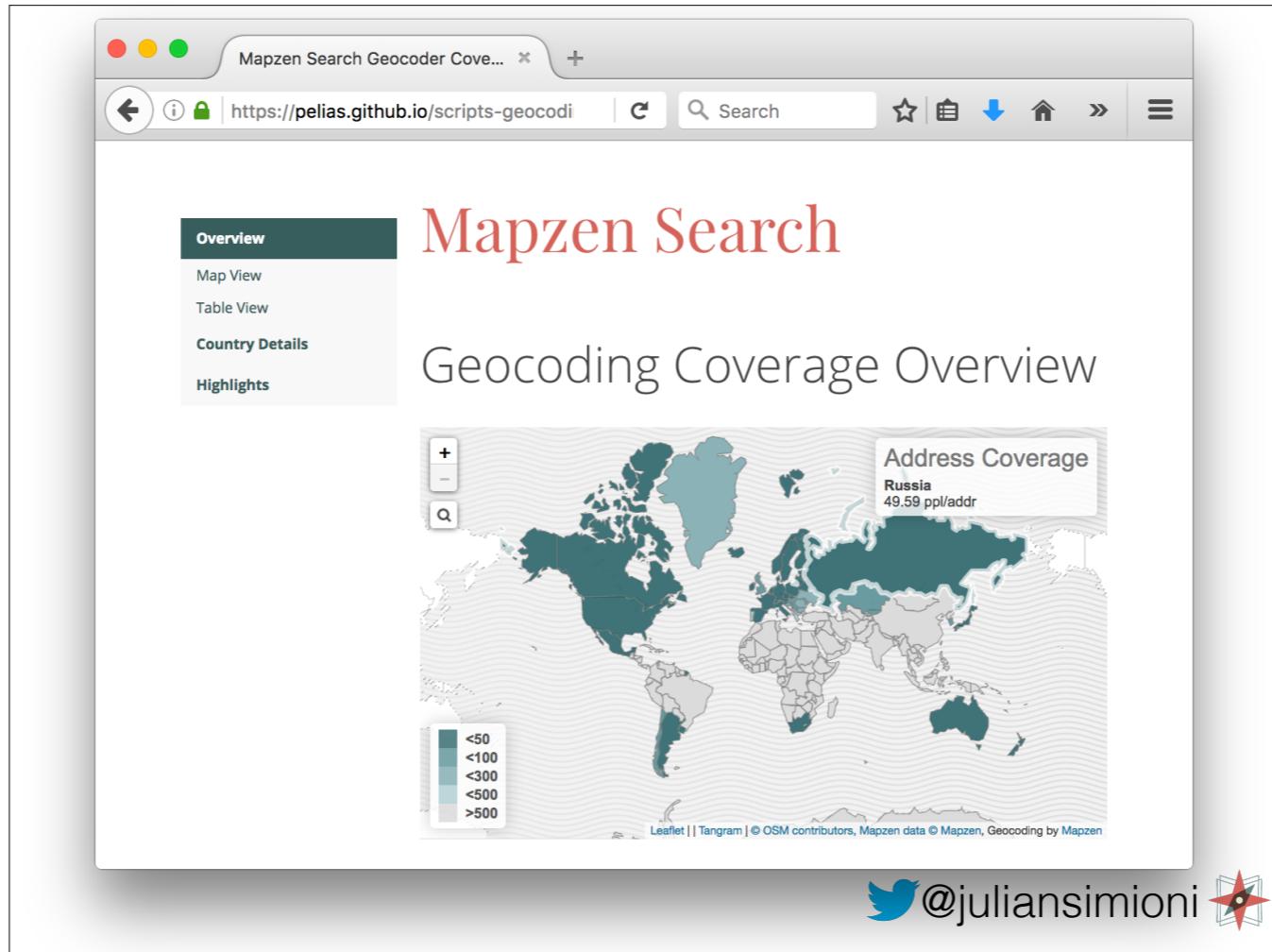
Sometimes the ways people search for places will be real weird.

So, indexing the planet.
How's that going?

@juliansimioni 



This is our Pelias build dashboard. It shows the progress of our planet builds and a summary of what they contain. We build every week and every week the numbers creep a little higher!



We also want to know more than just the totals, so we built a tool to analyze our coverage in every country.

Mapzen Search Geocoder Coverage

https://pelias.github.io/scripts-geocoding-coverage/

Overview

Map View

Table View

Country Details

Highlights

Country	Population	Total Records	People per Addr	People per Venue
Afghanistan (AFG)	29121286	76430	7280321.50	676.72
Albania (ALB)	2986952	13956	3564.38	355.93
Algeria (DZA)	34586184	89401	34759.98	450.95
Andorra (AND)	84000	4903	158.19	21.36
Angola (AGO)	13068161	54111	20612.24	351.72
Antigua and Barbuda (ATG)	86754	1294	6673.38	77.39
Argentina (ARG)	41343201	1122235	43.04	273.46
Armenia (ARM)	2968000	23676	338.97	226.00
Aruba (ABW)	71566	2471	102.09	42.70
ATA	unknown	12515	unknown	unknown
Australia (AUS)	21515754	19923933	1.10	54.56
Austria (AUT)	8205000	3732201	2.41	26.09
Azerbaijan (AZE)	8303512	22186	1750.32	663.70
Bahrain (BHR)	738004	2165	56769.54	376.15
Bangladesh (BGD)	156118464	83175	13171.22	6815.02
Barbados (BRB)	285653	3336	398.96	136.55

<https://pelias.github.io/scripts-geocoding-coverage/>

@juliansimioni 

Screenshot of the HOT Tasking Manager website showing a project for Hurricane Matthew.

The page title is "HOT Tasking Manager" and the URL is "tasks.hotosm.org".

The main heading is "Projects".

A search bar and a sorting dropdown ("Sort by: High priority first") are present.

The project details are as follows:

- #2205 Hurricane Matthew: Nippes coast**
- Current #HurricaneMatthew Project**
- Author: HOT**
- Requesting organization: Red Cross**
- Priority: Urgent**

Buildings Only

Hurricane Matthew a category 4 hurricane caused catastrophic damage to western Haiti. Over 2 feet of rain fell, causing flooding, damaging homes and washing away infrastructure. According to reports over 15,000 people have been displaced including destroying almost 2,000 homes. Bridges and roads have been affected including critical links for relief efforts.

This data will be used for a detailed damage assessment, relief planning, and for future cholera prevention efforts.

Created by [dkunce](#) - Updated less than a minute ago - Priority: urgent

A small world map shows the location of the tasking area.

At the bottom right of the screenshot is a Twitter handle [@juliansimioni](#) accompanied by a small red and white map icon.

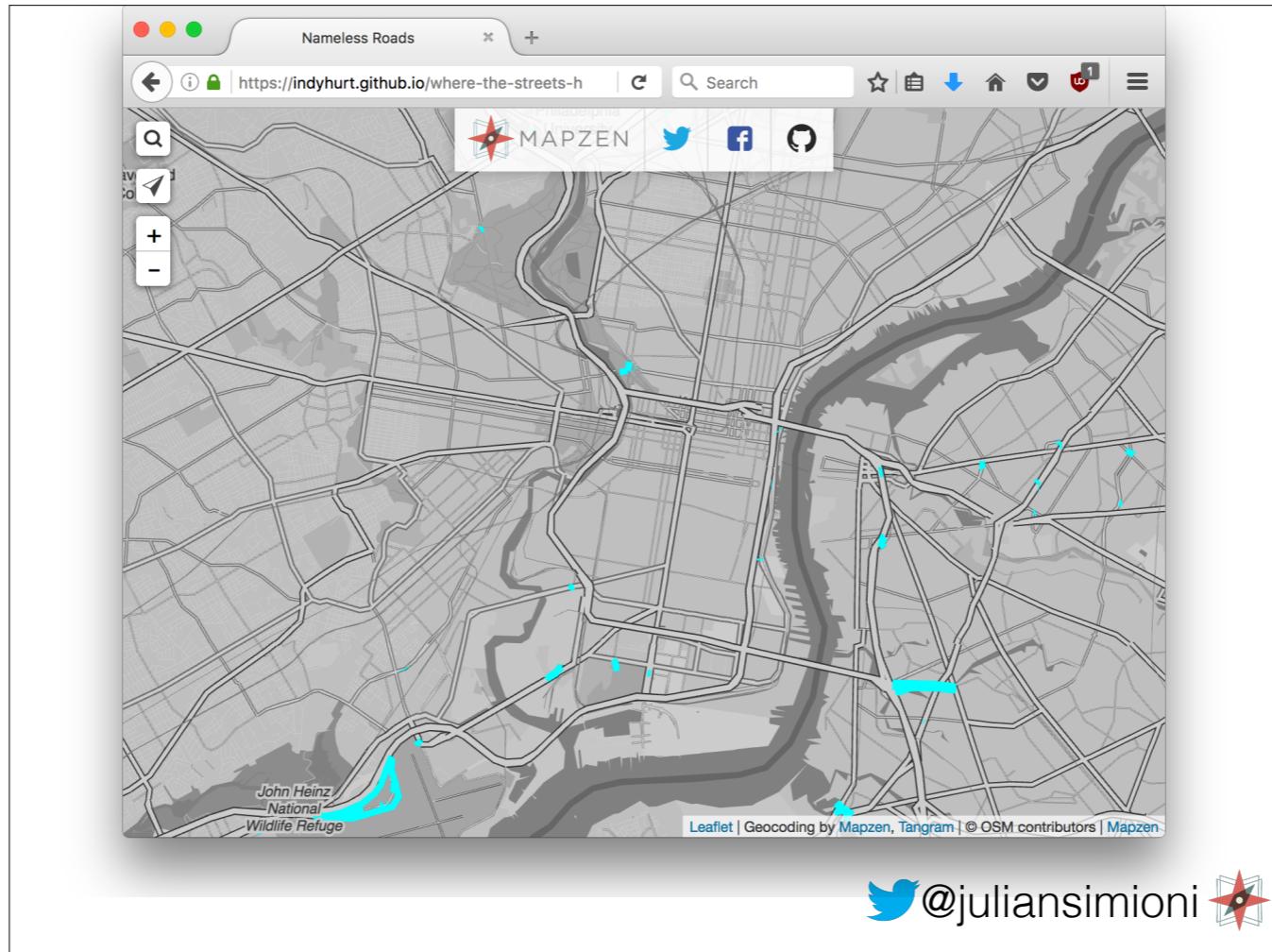
Since Pelias uses only open data, we rely partially on the community to help add more data. Of course adding data to projects like OpenStreetMap and OpenAddresses can also have uses beyond geocoding, and with the recent Hurricane Matthew, contributing via projects like Humanitarian OpenStreetMap is a great idea.

The screenshot shows a blog post on the Mapzen website. The header includes the Mapzen logo, navigation links for PRODUCTS, DOCUMENTATION, BLOG (which is underlined in red), and DEVELOPERS. Below the header is a back-link button labeled '< BLOG'. The main title of the post is 'Targeted Editing – Where the streets have no names...'. Underneath the title are two small gray rectangular tags with the text 'targeted-editing' and 'osm'. The post content begins with the question 'Should all streets have names? Actually, no...'. It then describes the 'Targeted Editing' series and its purpose. The URL of the post is displayed below the title.

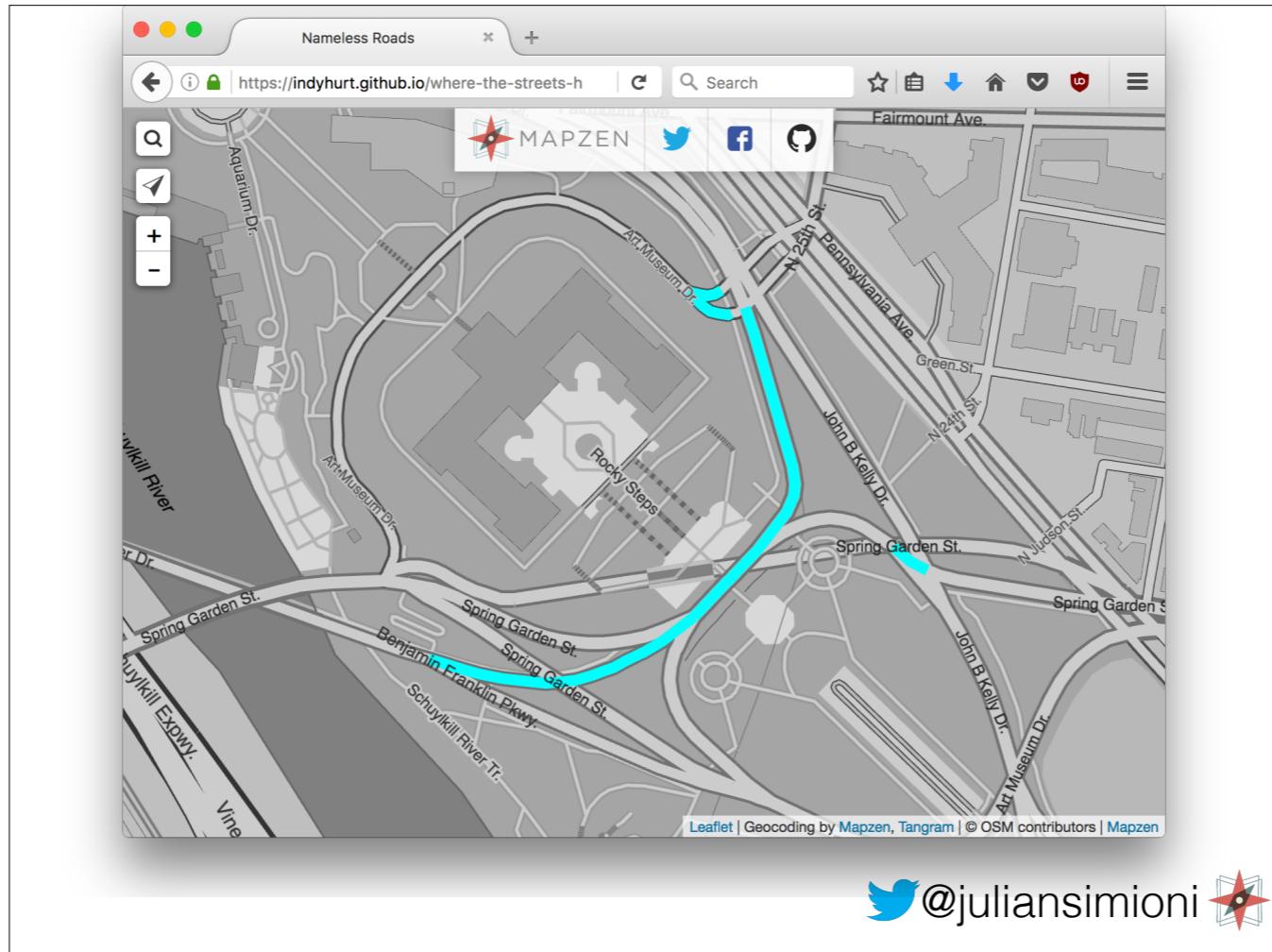
<https://mapzen.com/blog/targeted-editing-no-name-roads/>

@juliansimioni 

At Mapzen we also work on tools to help people find other ways to contribute. Combined with tracing roads and buildings from satellite photos, recording the names of roads and places is also important.



You're almost done Philly! But a few places are missing road names.



@juliansimioni

This one feels like an easy one to fix, hopefully someone will do it. Also note that the Rocky Steps, of course, are already named.

Peter Brodersen @peterbrodersen · Apr 21
Dear @mapillary. You are playing with fire.

STREET LEVEL PHOTOS

Discover places and capture the world around you.
Join us and help visualize the world.

Copenhagen

SEARCH

Copenhagen, København, Sweden

Copenhagen, NY, USA

Village of Copenhagen, Copenhagen, NY, USA

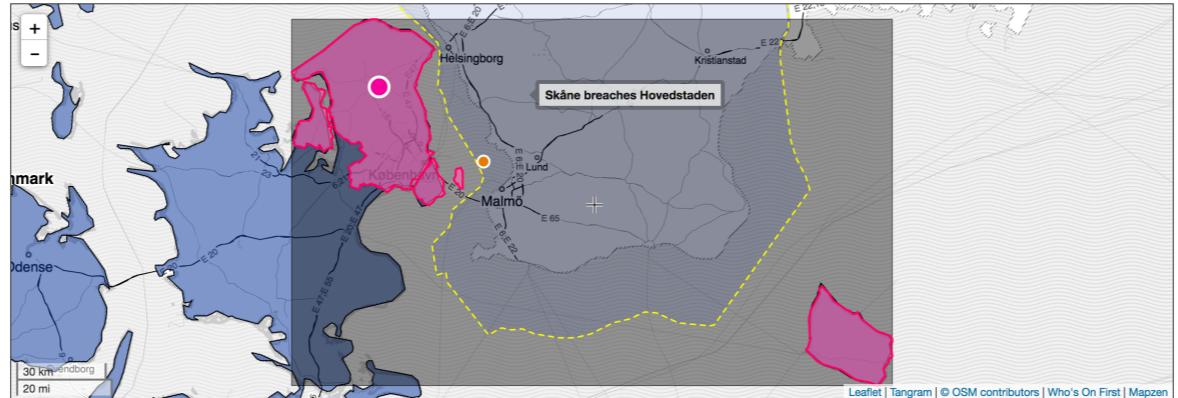
Copenhagen, Crawford, NE, USA

9 Copenhagen Court, Colonie, NY USA

@juliansimioni 

It's not just up to the community, we have our share of issues to work on. One time we had a bug where we told everyone Copenhagen was in Sweden. It's in Denmark and since the two countries have been to war a few times, this is a bit of a touchy subject. Oops.

Hovedstaden 856/825/81/85682581.geojson



[https://mapzen.com/blog/
assault-on-copenhagen/](https://mapzen.com/blog/assault-on-copenhagen/)

@juliansimioni 

Here's the full story there, definitely take a few minutes to read it and learn a lot about history and geometry



And how's the searching going?

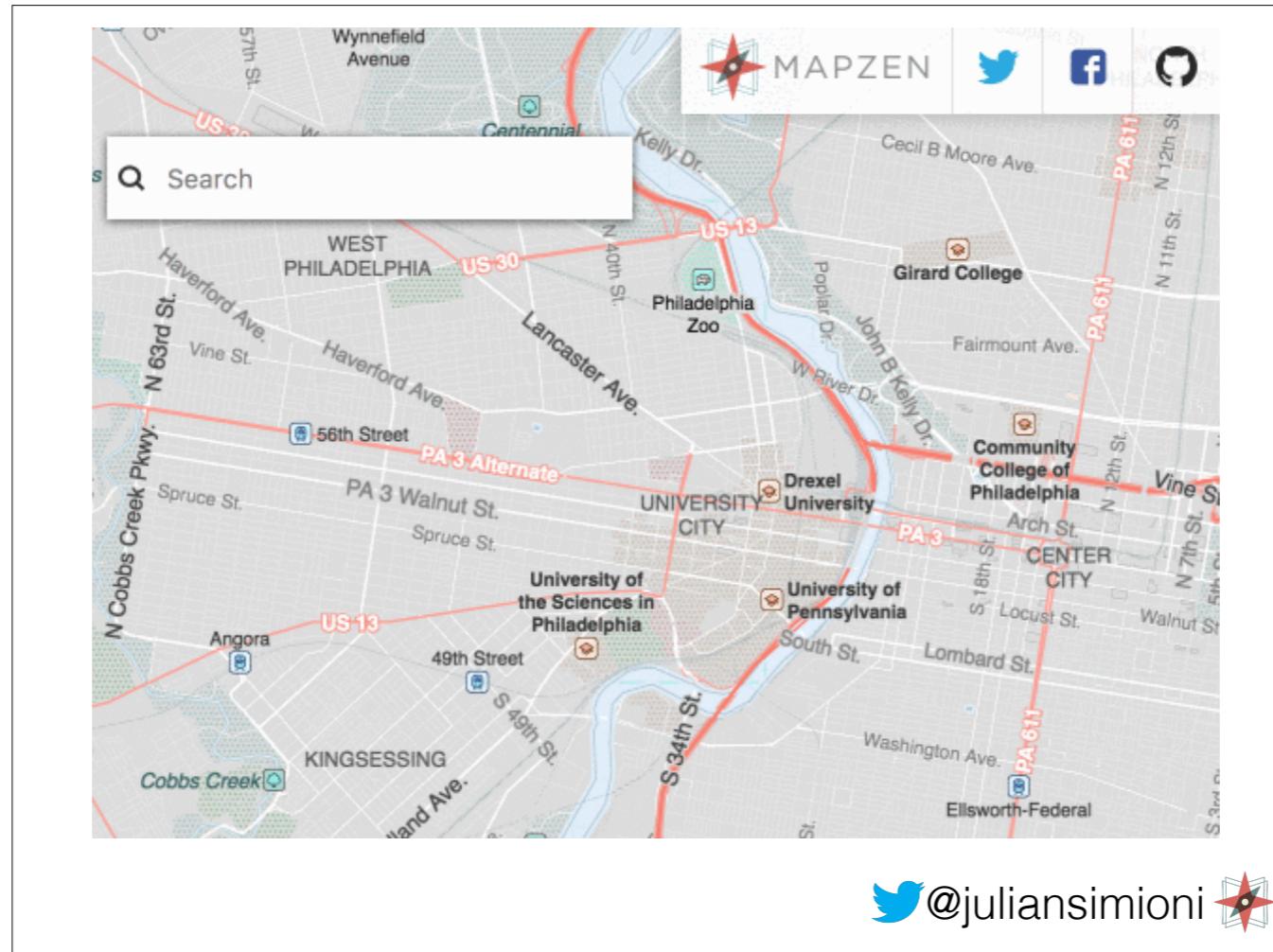
Shout out to all the Hall & Oates fans

 @juliansimioni 

Lets talk about Abbrevs.

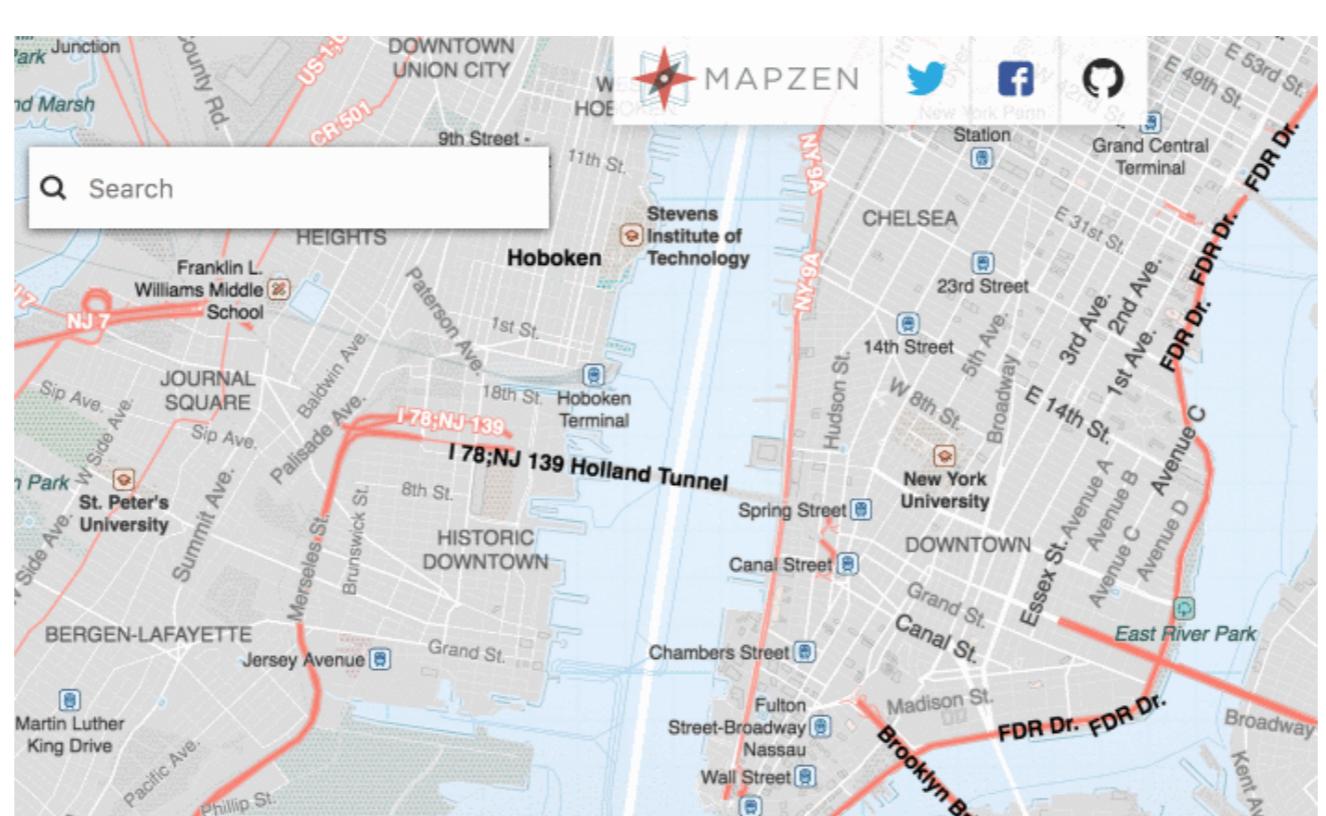
 @juliansimioni 

There are many reasons why searching through a database of the whole world is hard, but abbreviations are a good example of how something “simple” can become a lot of trouble.



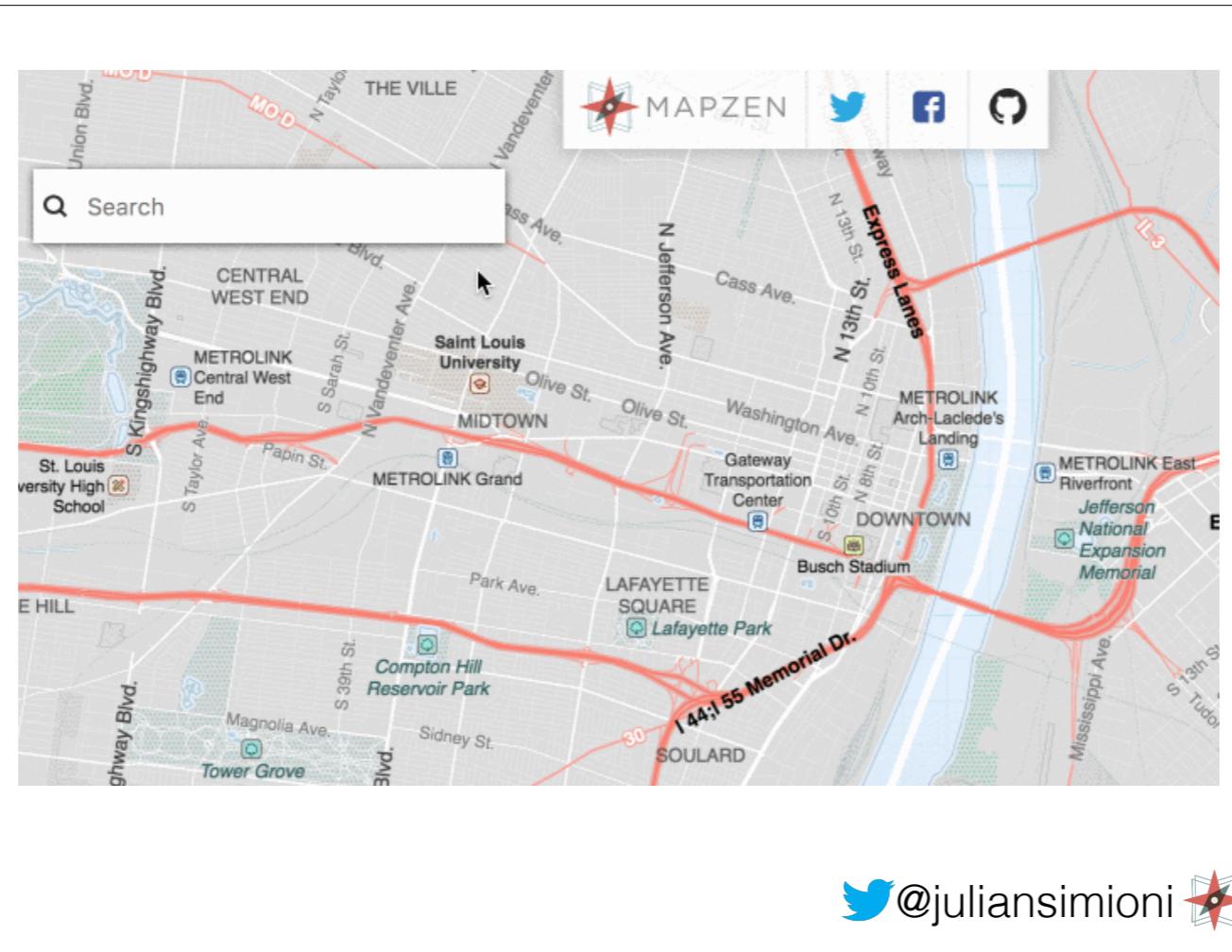
Sometimes, it's obvious what an abbreviation means. We have no trouble matching "Market Street" when you type "Market St".

But “St” doesn’t always mean Street...



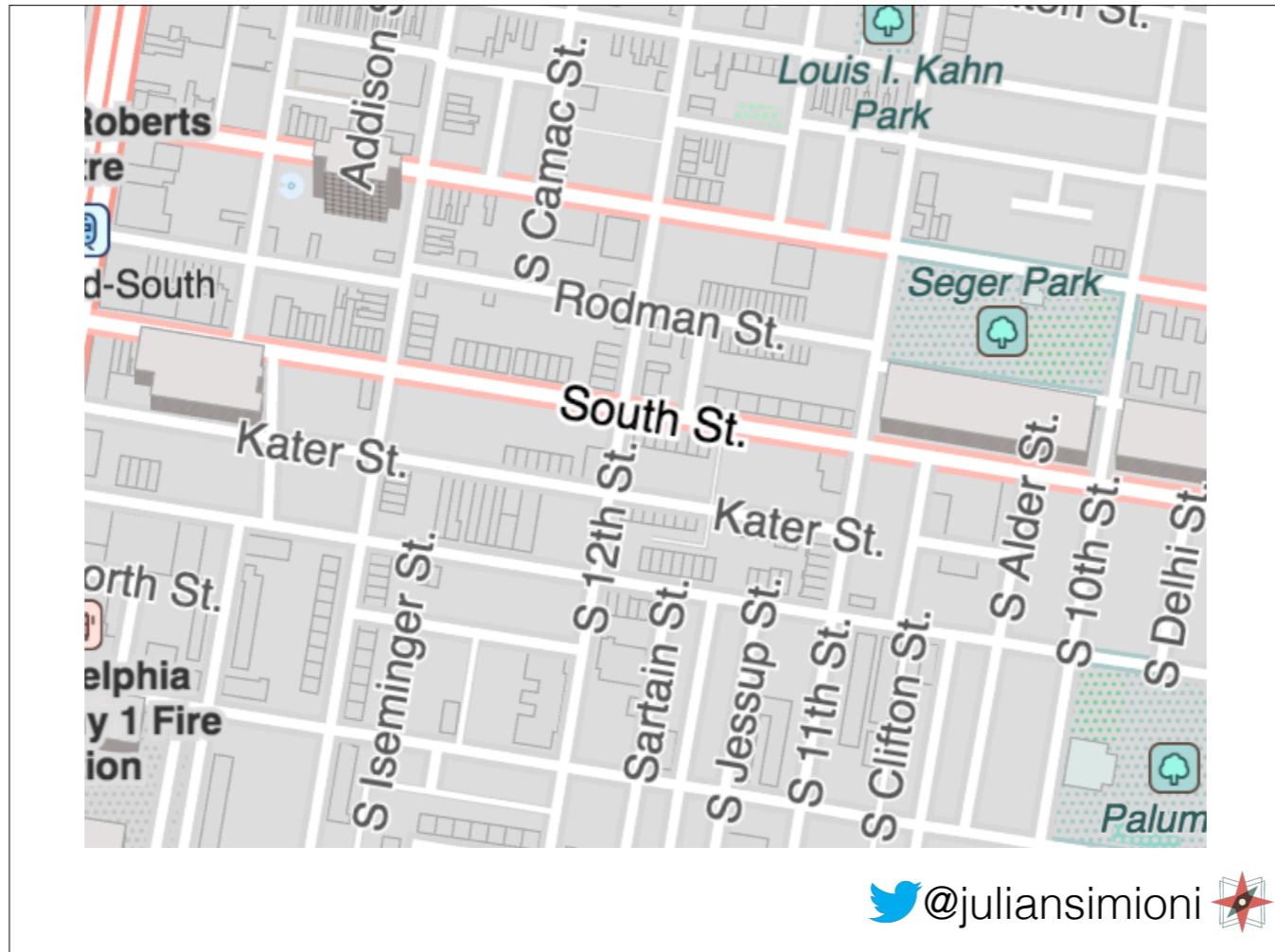
@juliansimioni 

Searching for Saint Marks Place in NYC works only if you spell out Saint. If you use “St”, you’re out of luck (for now, we’ll fix this soon!).



@juliansimioni

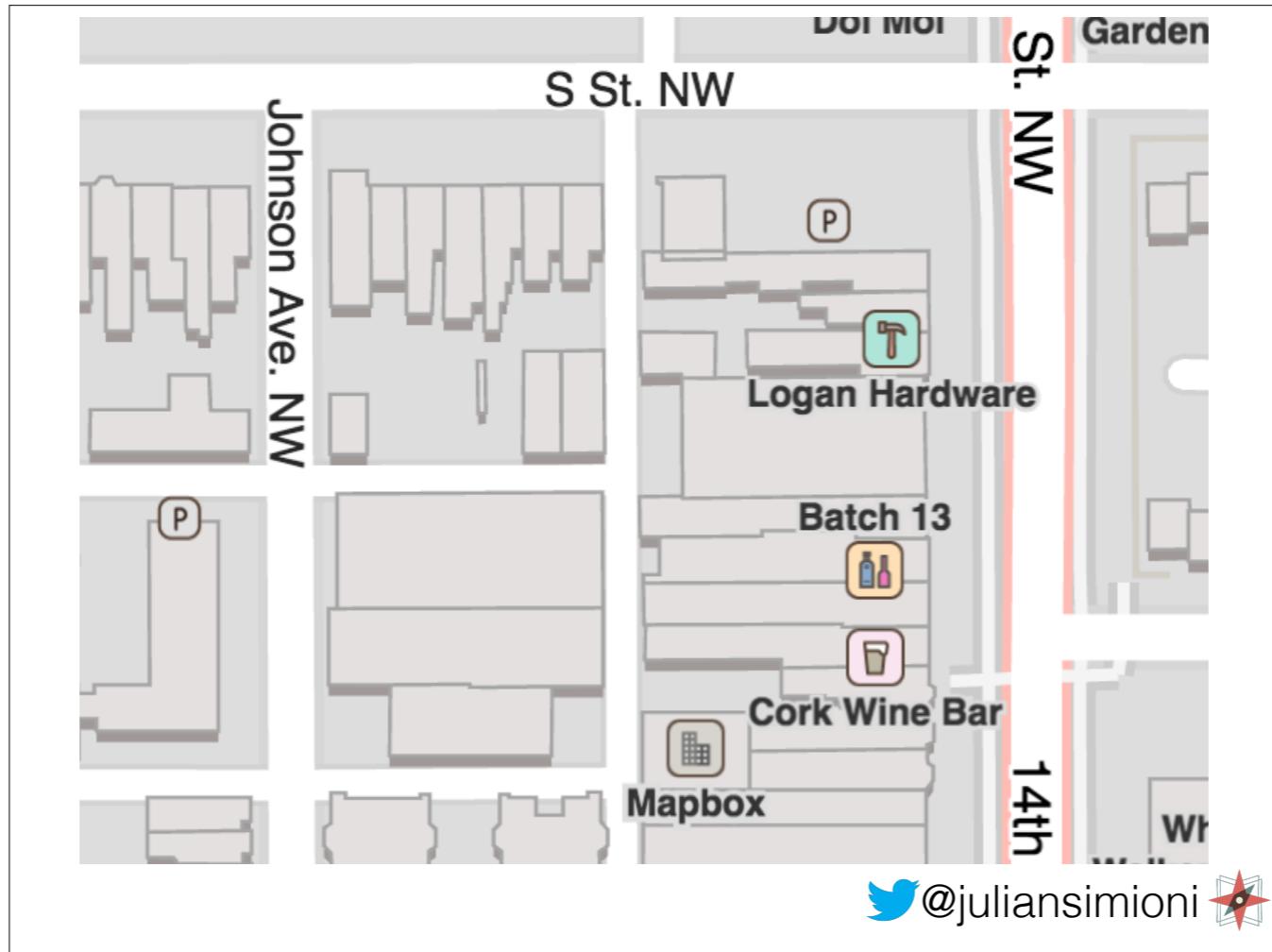
Sometimes “St” can mean two things in the same address. Residents of St. Louis are probably all too familiar with this.



@juliansimioni



A lot of words that feel like they are unambiguous when abbreviated actually aren't. I'm sure someone occasionally writes South St. as "S St", and the world probably carries on as normal.



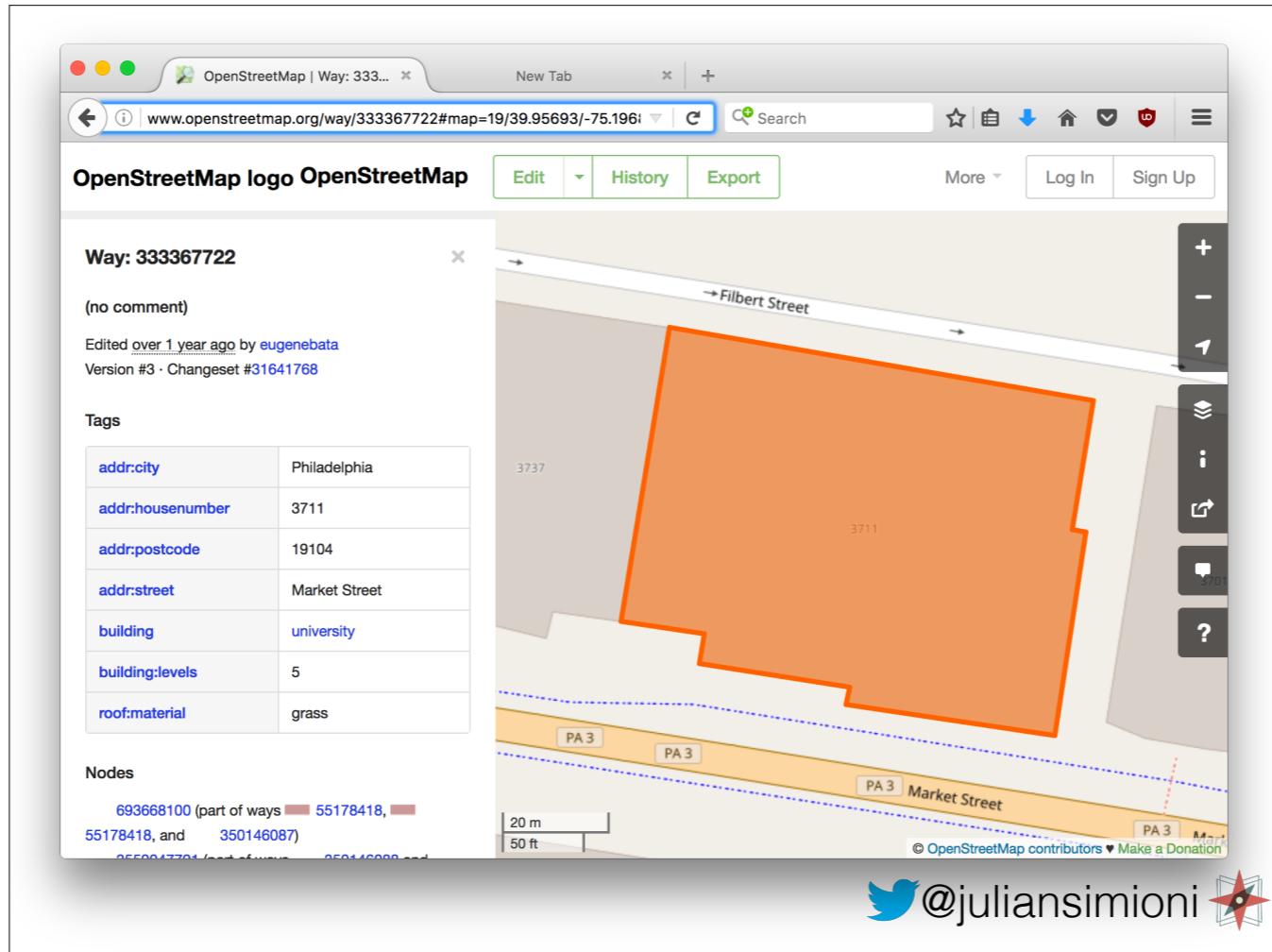
But that won't work once you get to Washington D.C. There's actually an S Street (like the letter) there!

Also, everyone wave to Mapbox HQ!!

```
4 module.exports = function(input, opts) {
5   // parse the base address
6   return parser(input, extend({
7     state: {
8       AL: /(^\alabama|^AL$)/i,
9       AK: /(^\alaska|^AK$)/i,
10      AS: /(^\american\ssamoa|^AS$)/i,
11      AZ: /(^\arizona|^AZ$)/i,
12      AR: /(^\arkansas|^AR$)/i,
13      CA: /(^\california|^CA$)/i,
14      CO: /(^\colorado|^CO$)/i,
15      CT: /(^\connecticut|^CT$)/i,
16      DE: /(^\delaware|^DE$)/i,
17      DC: /(^\district\sof\scolumbia|^DC$)/i,
```



The most straightforward approach to geocoding is to try to program in all the edge cases. Anyone who's worked with regular expressions knows how this ends.



But there already is, in a sense, a database of all the edge cases, and it's called the data we're already using.

Way: 333367722

X

(no comment)

Edited over 1 year ago by eugenebata

Version #3 · Changeset #31641768

Tags

addr:city	Philadelphia
addr:housenumber	3711
addr:postcode	19104
addr:street	Market Street
building	university
building:levels	5
roof:material	grass

 @juliansimioni 

Is the roof really grass?

Aren't we on the 8th floor?

Can we use this data?



Machine learning is like a deep-fat fryer. If you've never
deep-fried something before, you think to yourself:
"This is amazing! I bet this would work on anything!"

http://idlewords.com/talks/deep_fried_data.htm

 @juliansimioni 

I just wanted to use this great joke

The screenshot shows the GitHub repository page for 'openvenues / libpostal'. The repository has 1,678 commits, 2 branches, 2 releases, and 14 contributors. It is an MIT-licensed C library for parsing/normalizing street addresses. The README.md file describes it as a C library for parsing/normalizing street addresses around the world using statistical NLP and open data. The repository has 49 issues, 1 pull request, 0 projects, 0 wiki pages, and 41 forks.

<https://github.com/openvenues/libpostal>

@juliansimioni

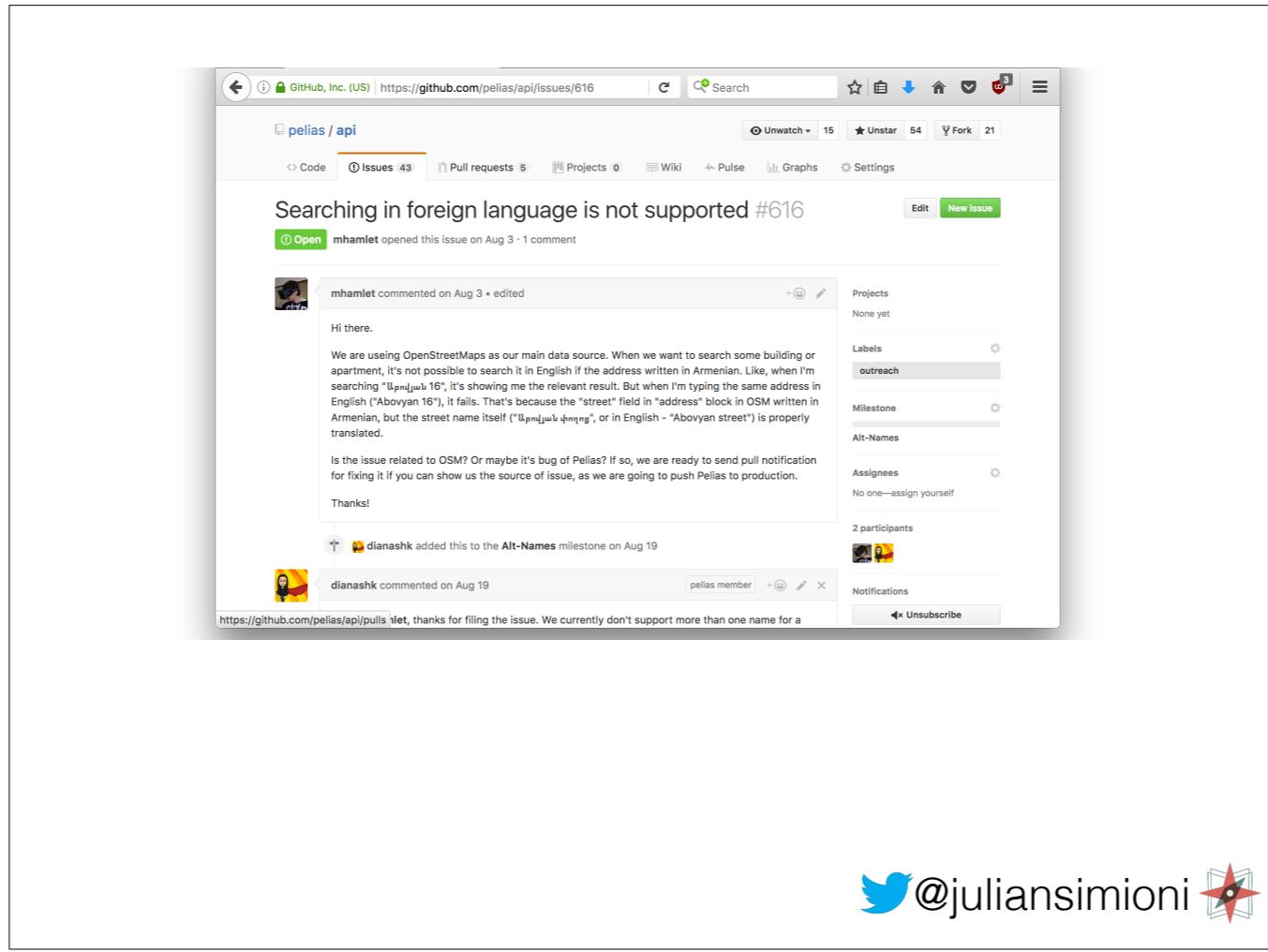
For the past year or two Mapzen has been sponsoring an incredible project to build a street address and place parser trained on OpenStreetMap data. It allows us to use machine learning to build the core of the geocoder: the input parser.

(It went live in Pelias a few days after this talk!)

What about the people actually writing the code?

@juliansimioni 

Okay, enough tech talk. We're also an open source project run by real humans and with lots of other humans helping out. We've learned a lot more about this and we want to continue to get better. Here's one story of something we learned.



@juliansimioni

Every open source project loves contributors, and thankfully lots of people are eager to contribute to Pelias. But sometimes there's a mismatch in expectations.

Is the issue related to OSM? Or maybe it's bug of Pelias?

If so, we are ready to send pull notification



Sometimes what seems like a bug to a user is actually related to an entire huge feature that we had planned to work on in a few months, because it's so big. But they don't know that. Unless...

<http://pelias.io>

Pelias

A modular, open-source geocoder built on top of ElasticSearch for fast geocoding.

chat on gitter

Pelias is the project used to power [Mapzen Search](#)

What's a geocoder do anyway?

Geocoding is the process of transforming input text, such as an address, or a name of a place—to a location on the earth's surface.

@juliansimioni 

We make a home page for the project that goes beyond just talking about the current state of the project, and also lists our roadmap. We also describe a rough outline of how we want to approach certain features and challenges, so that everyone's on the same page. It's been a big help, and we'll continue to make it even easier for people to know how, why, and when to contribute to Pelias.