

Geocoding in the Open

04-27-2016 // Diana Shkolnikov



Agenda

→ Mapzen

- *Mission*
- *Products*

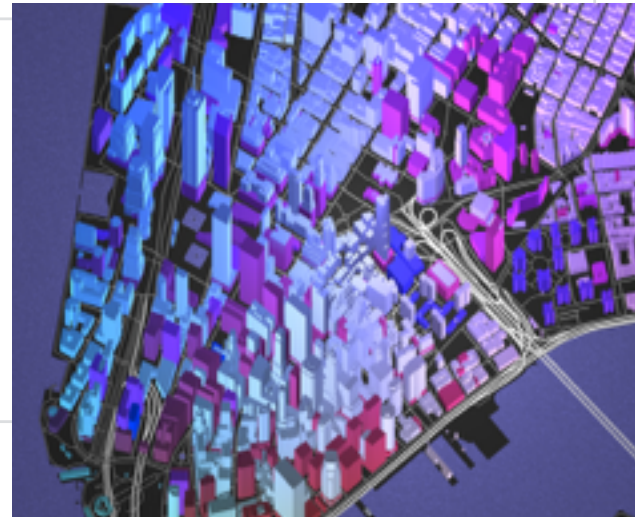
→ Geocoding

- *Key features*
- *Pelias engine*

→ Open Data

- *OpenStreetMap*
- *OpenAddresses*
- *Who's on First*

→ Demos: ooh, shiny!



MAPZEN

an open-source mapping lab

*builds open-source mapping tools and
collaborates on open geodata initiatives*

- we believe in the power of open
- communities should own the data that represents them
- communities should have direct free access to the tools that power their operations
- creating open source tools that make open data accessible, will encourage communities to create more open data
 - self-sustaining ecosystem

The screenshot shows a Google Maps interface with a search bar containing 'San Francisco'. Below the search bar, a list of results is displayed, including 'San Francisco, CA, USA', 'City of San Francisco, San Francisco, CA, USA', 'City and County of San Francisco, CA, USA', 'San Francisco Terepates, Mexico', 'San Francisco de Compostela, Mexico', 'San Francisco de Macoris, Dominican Republic', 'South San Francisco, CA, USA', 'San Francisco, Municipio Pánuco de Saragosa, Mexico', 'San Francisco, Argentina', and 'San Francisco, Costa Rica'. To the right of the list, a JSON data preview is shown, displaying the structure of the search results, including the name, location, and other details of the selected result.

Turn-by-Turn



various modes: car, pedestrian, bike
elevation service
time distance matrix

An aerial photograph of a city area, likely downtown Los Angeles, with a 3D model overlay. The model shows building footprints and heights, with a color-coded legend at the bottom right indicating different levels of development or density. The legend includes categories like 'REFILL', 'CINNABAR', and 'ZINC'.



[illegible]

Mobile SDKs



Mobile SDK wrappers for routing and search



First, let's quickly go over the key geocoder features



forward geocoding. find coordinates for an address

new yo

SEARCH

New York, NY

Pennsylvania Station New York, NY

New York University, New York, NY

New York Penn Station, New York, NY

autocomplete, aka typeahead

⌵

your current location

SEARCH

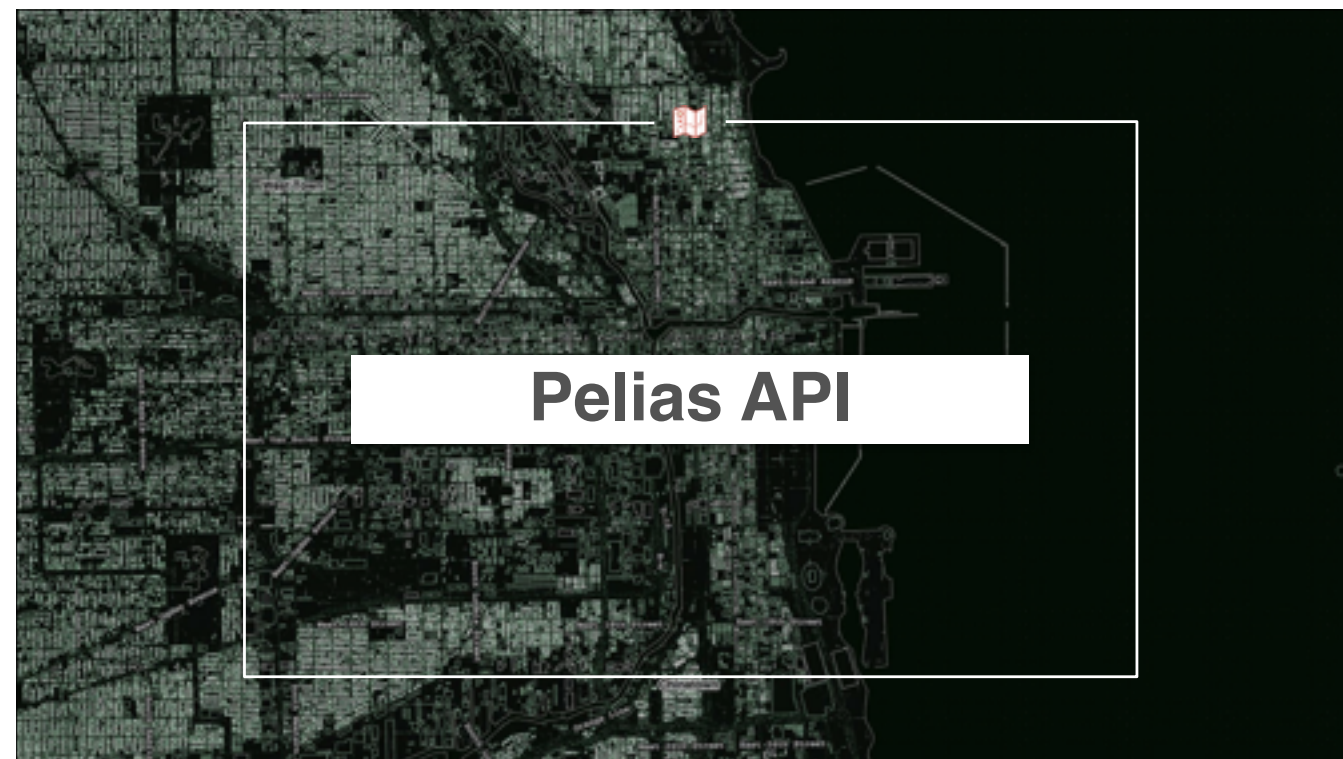
30 West 26th Street, Manhattan, NY

Hill Country Barbecue Market, Manhattan, NY

Samsung Accelerator, Manhattan, NY

Hanjan, Manhattan, NY

reverse geocoding: what's here?



Endpoints

/autocomplete

/search

/reverse

/place

/autocomplete

api_key: free authorization key

text: input query text

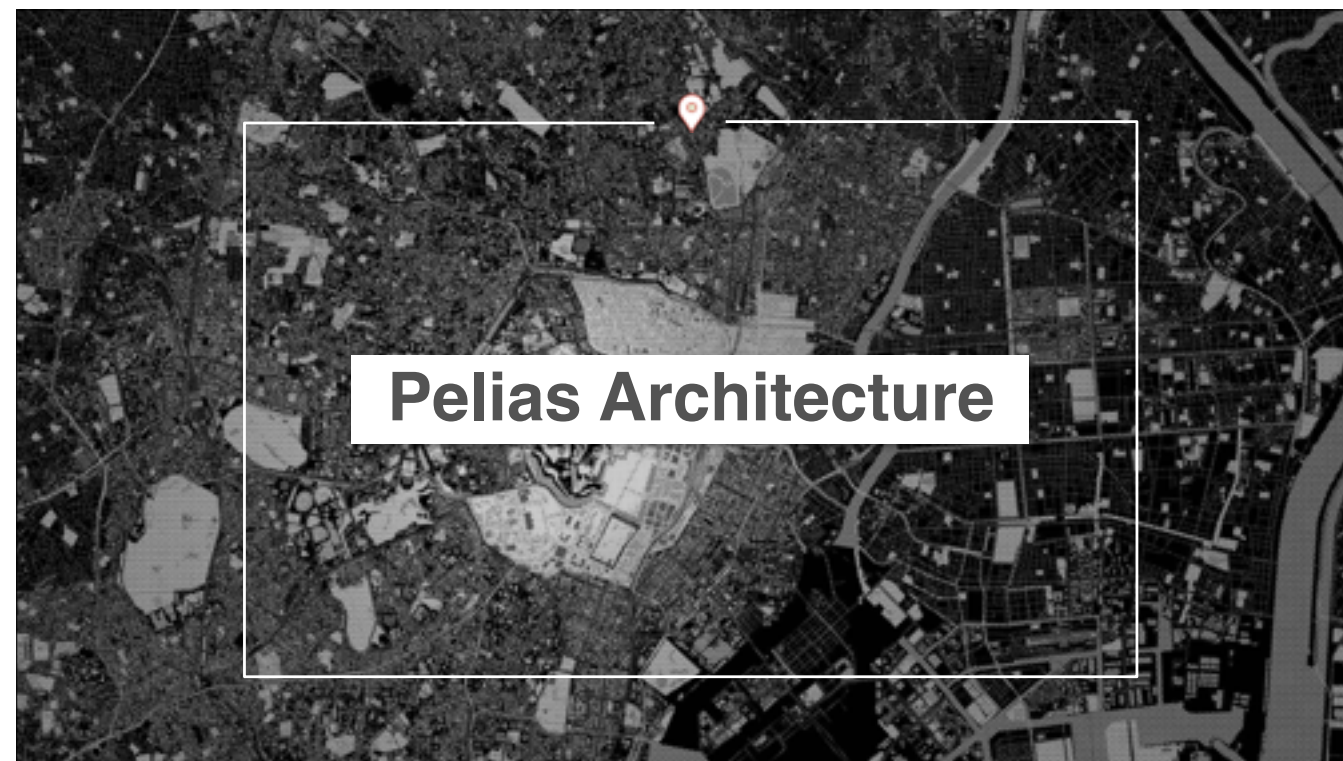
sources: openstreetmap, openaddresses, geonames, whosonfirst

layers: address,venue,neighbourhood,locality,borough,
localadmin,county,macrocounty,region,
marcoregion,country,coarse

focus.point.lat, focus.point.on:
geolocation for focusing results

/search

api_key: free authorization key
text: input query text
size: number of results to return
sources: openstreetmap, openaddresses, geonames, whosonfirst
layers: address,venue,neighbourhood,locality,borough,
localadmin,county,macrocounty,region,
marcoregion,country,coarse
focus.point.lat, focus.point.on:
geolocation for focusing results
boundary.country
boundary.circle [lat,lon,radius]
boundary.rect [min_lat, min_lon, max_lat, max_lon]
hard filters for limiting the scope of the search



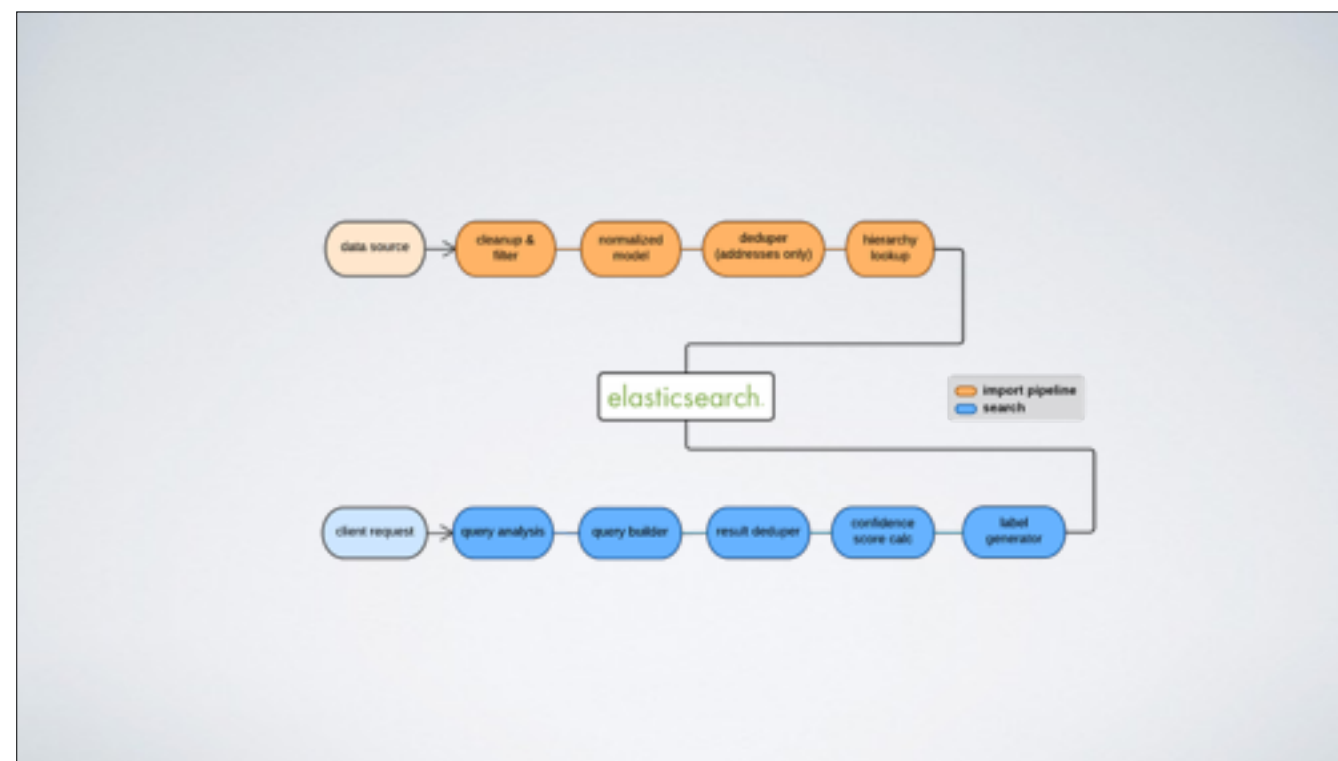
Pelias Architecture

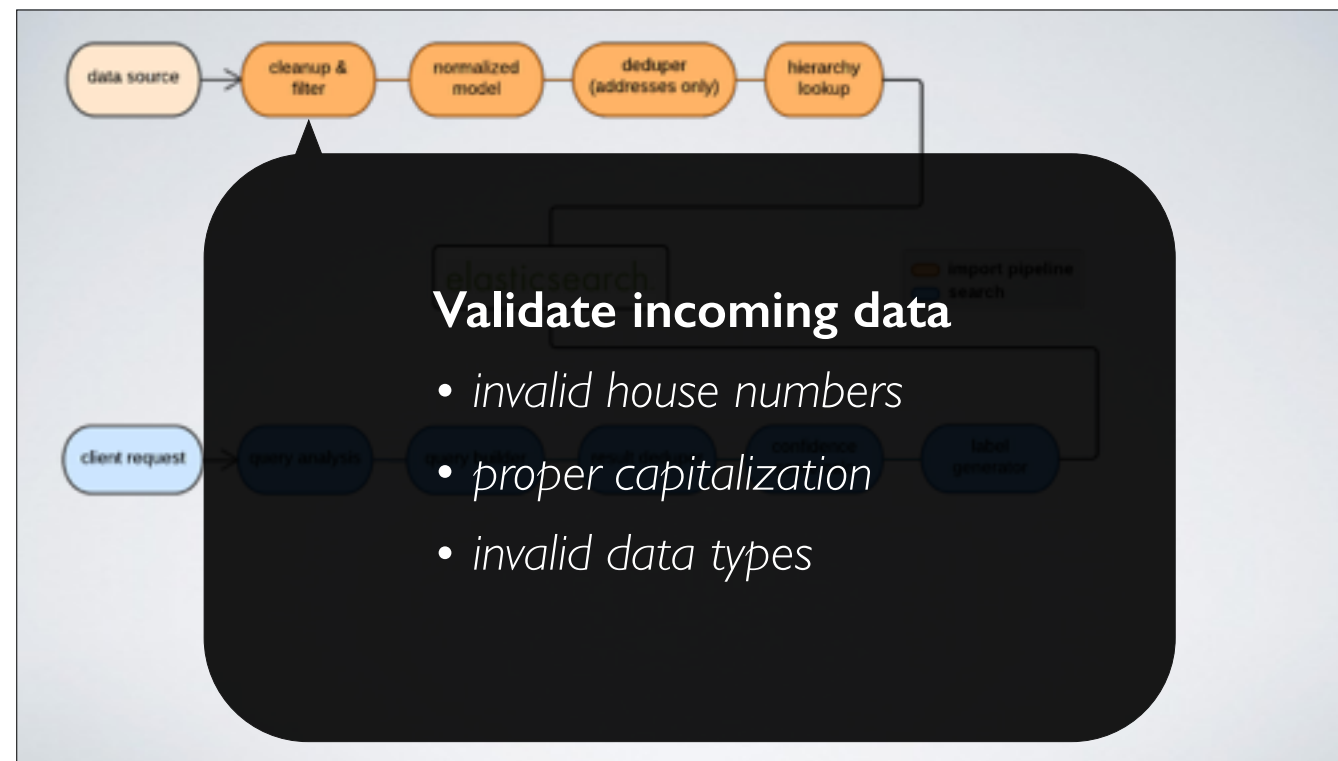


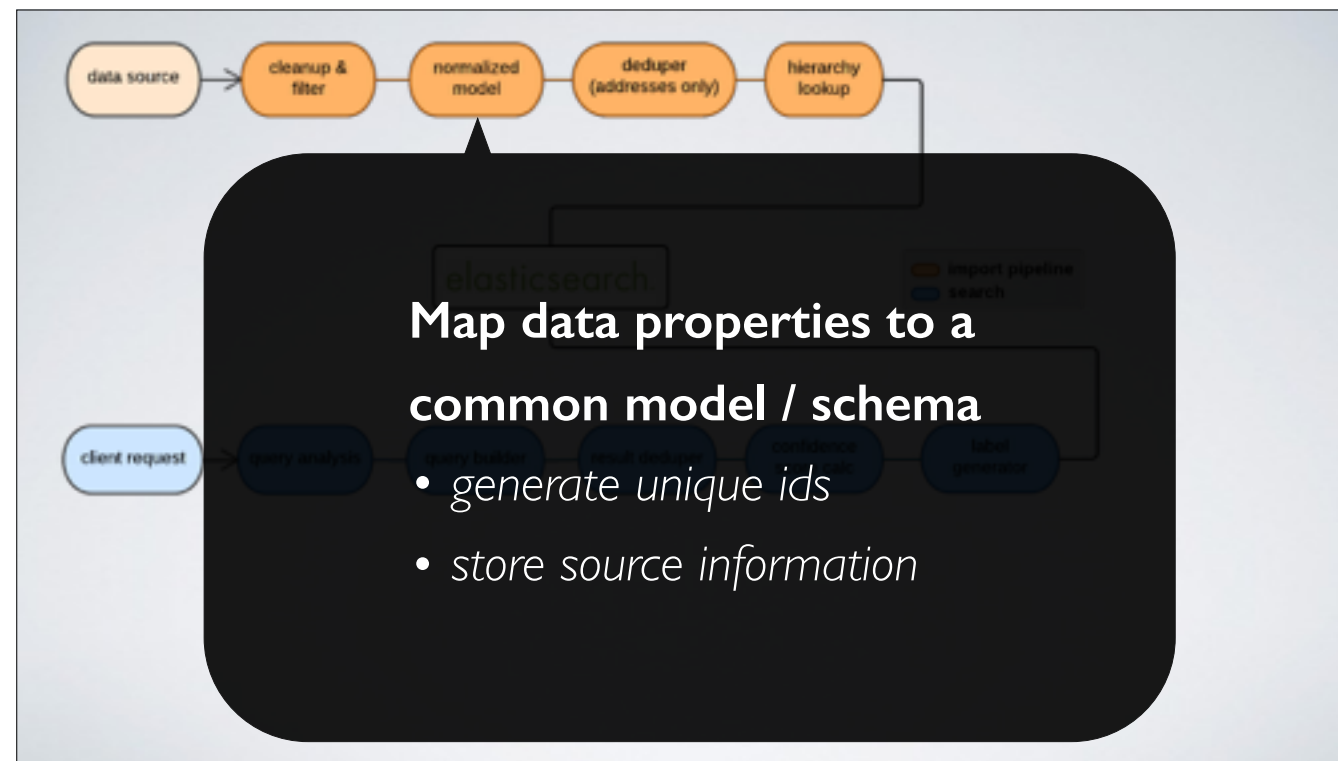
Elasticsearch

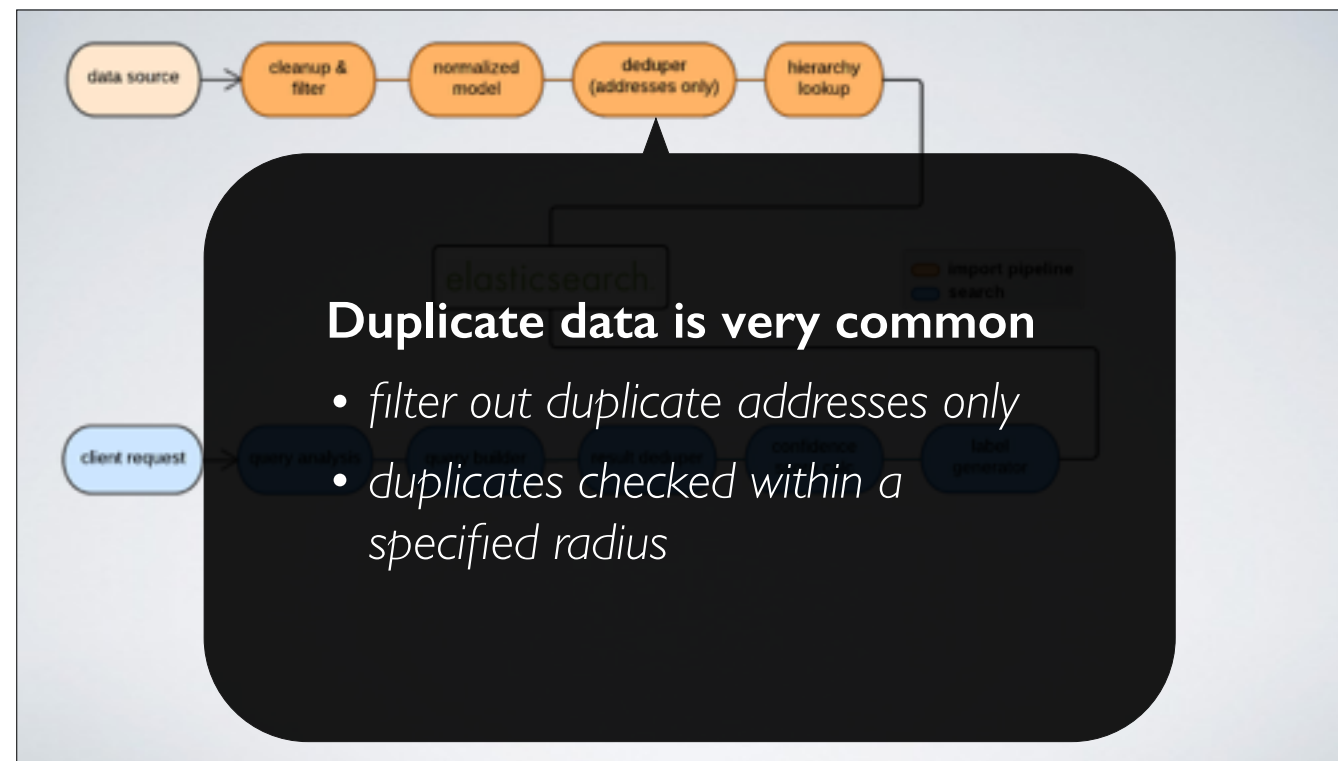


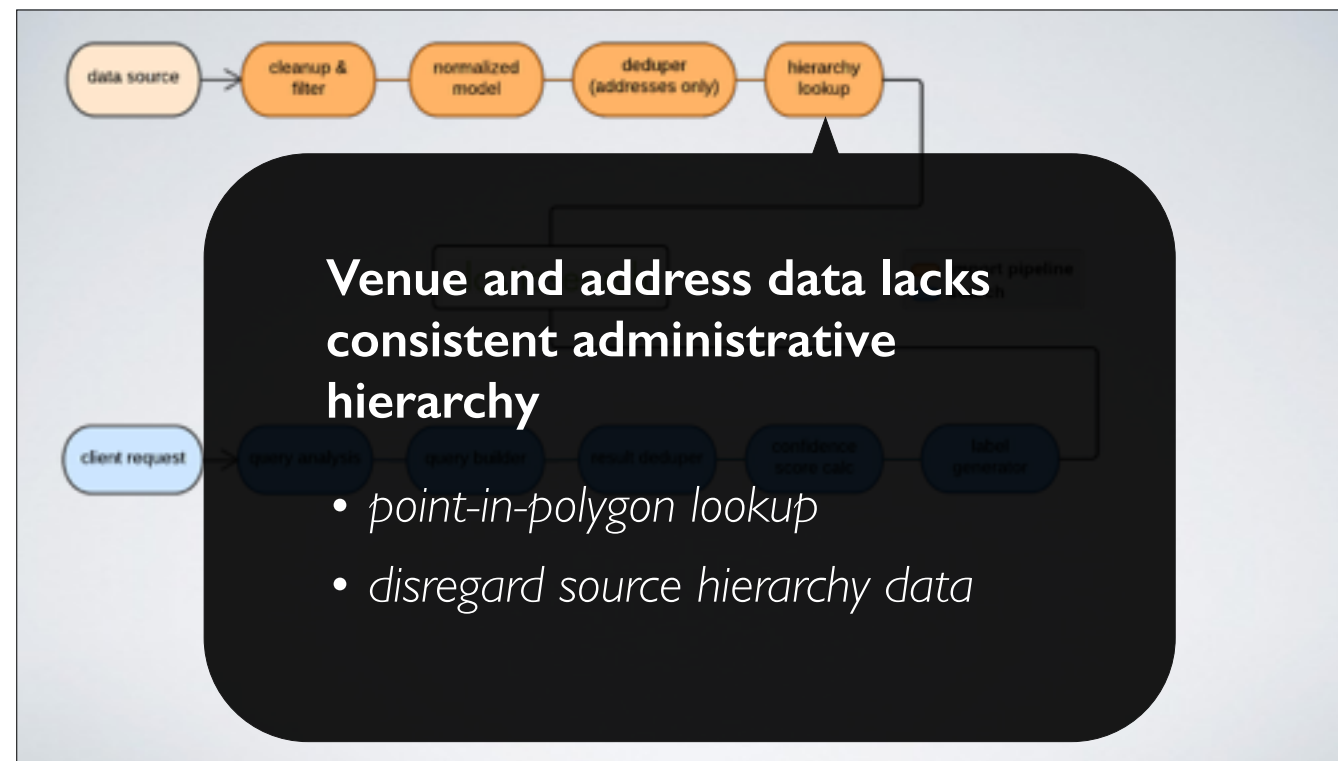
node.js



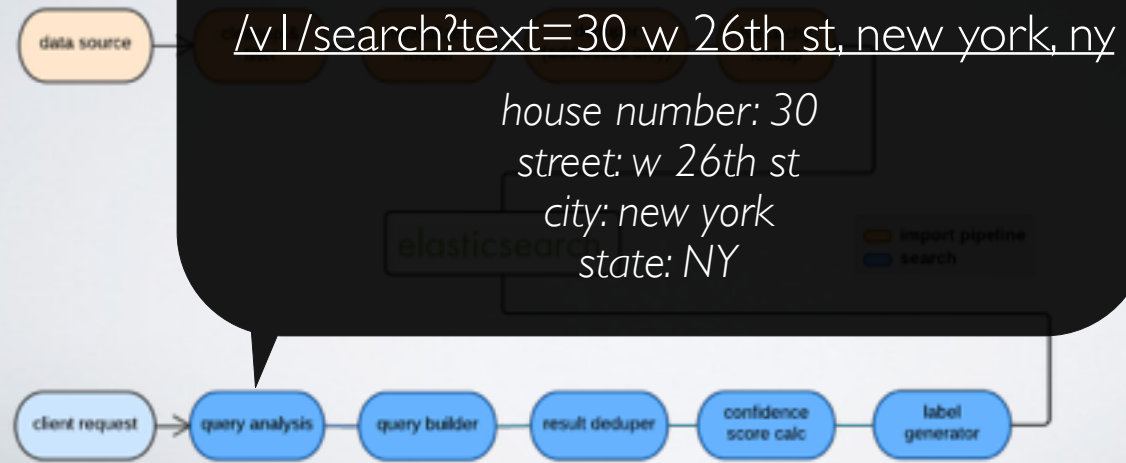








Parse and analyze the incoming query parameters



Generate the Elasticsearch query

- *add proper boosting*
- *insert match clauses for each address part*
- *add geographic focus and filters*

elasticsearch

import pipeline
search

data source

hierarchy
lookup

client request

query analysis

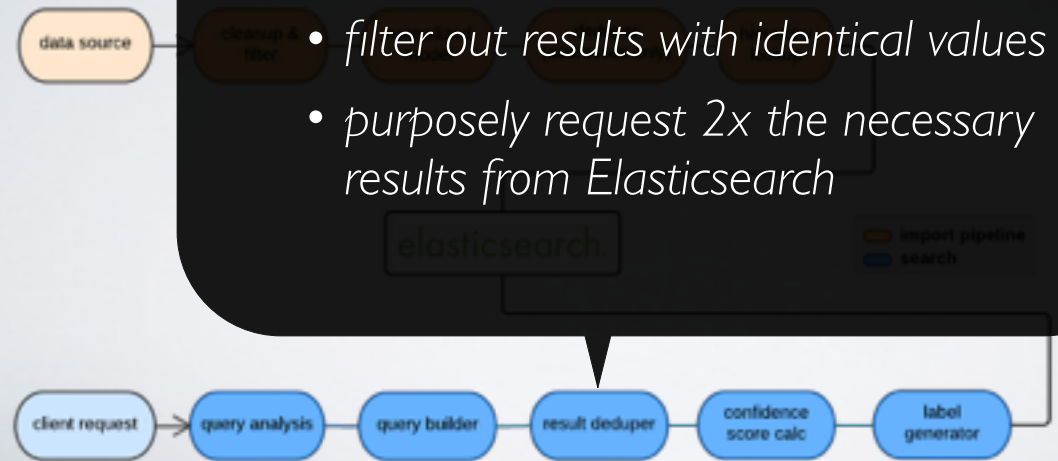
query builder

result deduper

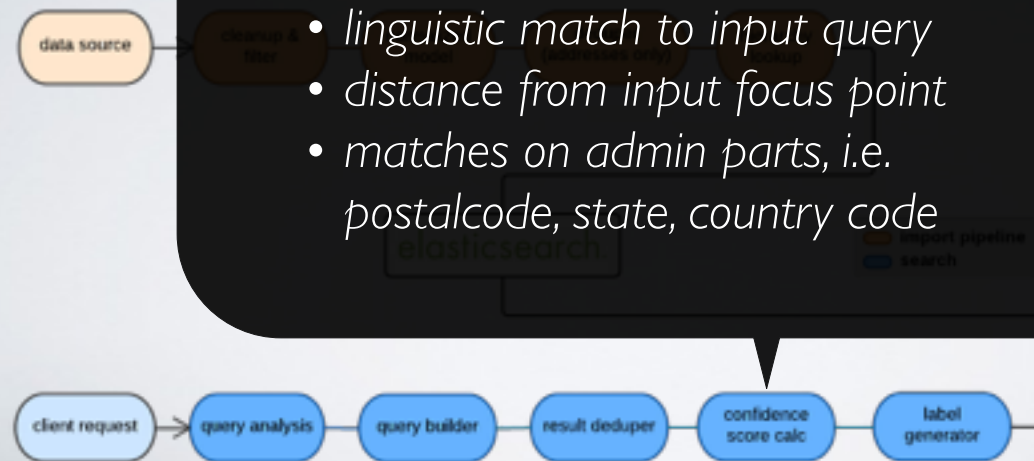
confidence
score calc

label
generator

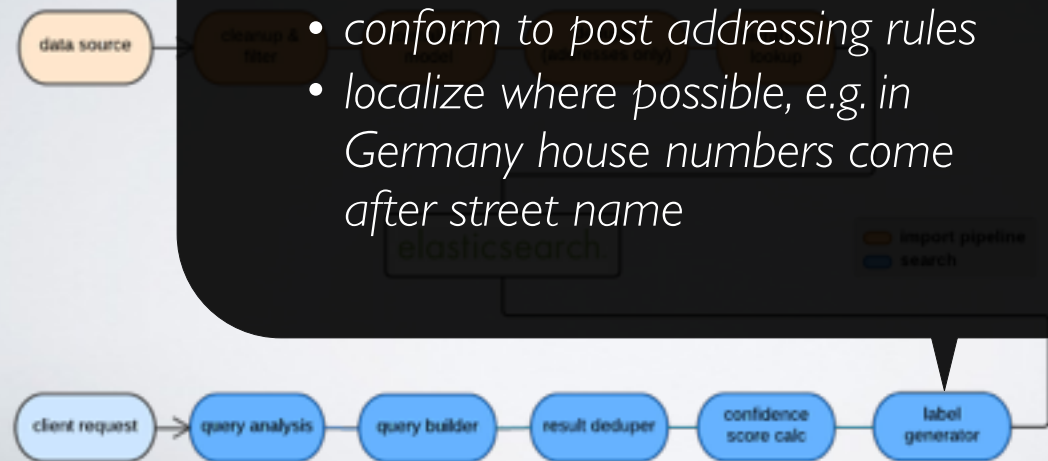
Combined data sources allow for duplicates in results



Compute the confidence score based on various factors



Generate result labels for easy presentation to end-user





Whoever controls the data, controls the public view of the truth.
Business gets marked as closed in Google and lose all their business because people stop going there. Business has no recourse for action and are at the mercy of Google change request process.
Having one or several large entities control something so critical to so many aspects of public safety is a scary proposition.

“

You are what you eat!

— Mom



Geocoders are a direct product of the data they ingest. Bad data cannot yield good search results!

correct / complete data
timely data updates
customizable data sources

Sadly, getting your hands on a reliable data set that doesn't cost a large fortune is virtually impossible.
We are banking on open data being the longterm solution. So we're here to build tools to make using open data awesome!

Open Data: WHY?

FREE

Accessible

Community maintained

Open Data: WHERE?



OpenStreetMap

openstreetmap.org

- Emphasizes LOCAL KNOWLEDGE
- Community driven: volunteers of all types contribute in different ways
- Discourage bulk/automated imports
- It's open under the ODBL license
 - use it freely anywhere with proper attribution
 - if you make changes to the data or build on top of it in certain ways you must make your resulting data available under the same license

Open Data: WHERE?

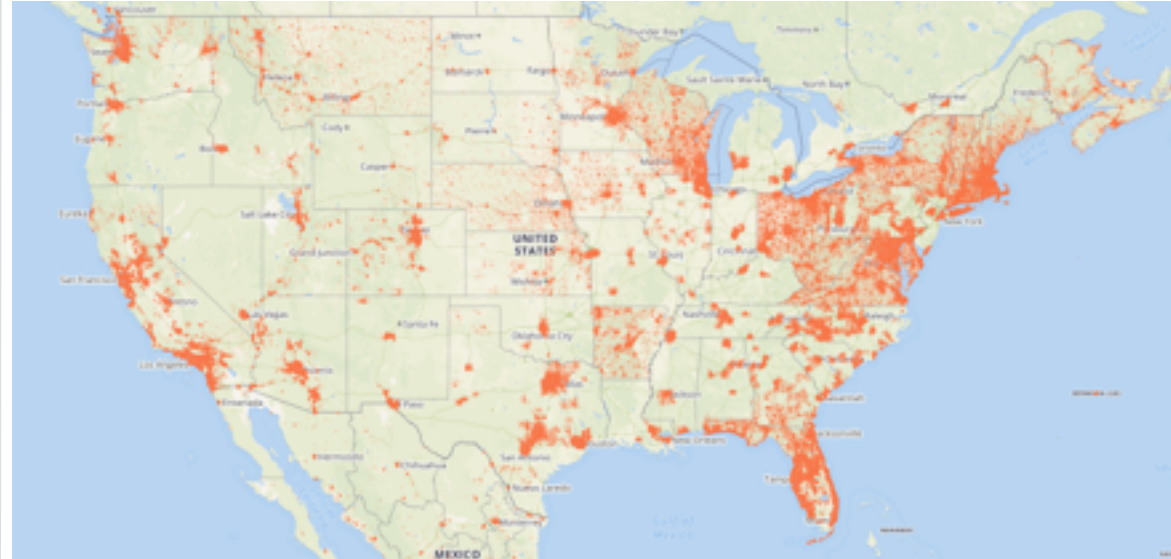


OpenAddresses

openaddresses.io

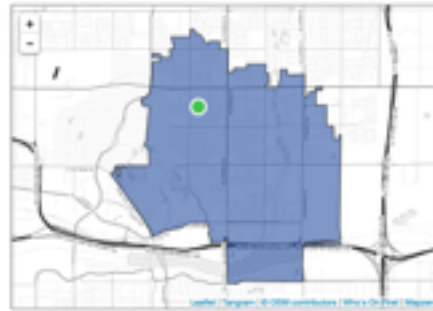
- Emphasizes government data
 - trusted reputable sources - counties / cities / states / countries
- Consists entirely of open local government data feeds from all over the world
- Entire countries have opened ALL their data, such as Australia and Mexico

OpenAddresses



- Coverage map isn't 100% up-to-date and new sources are constantly being added
- this gives you an idea of the coverage

Open Data: WHERE?

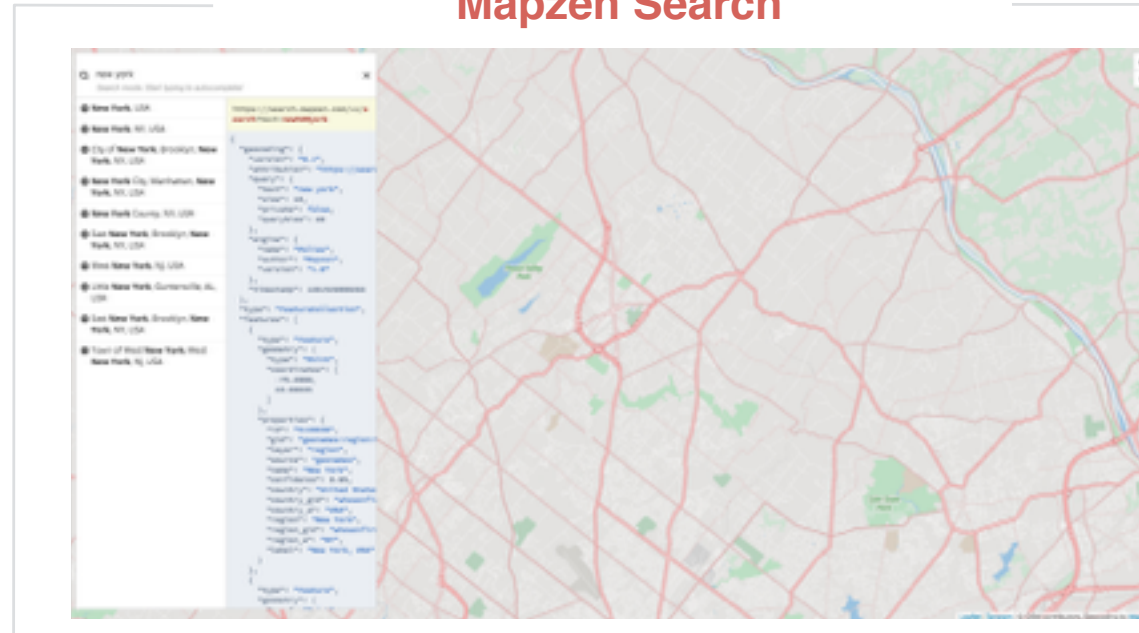


Who's On First

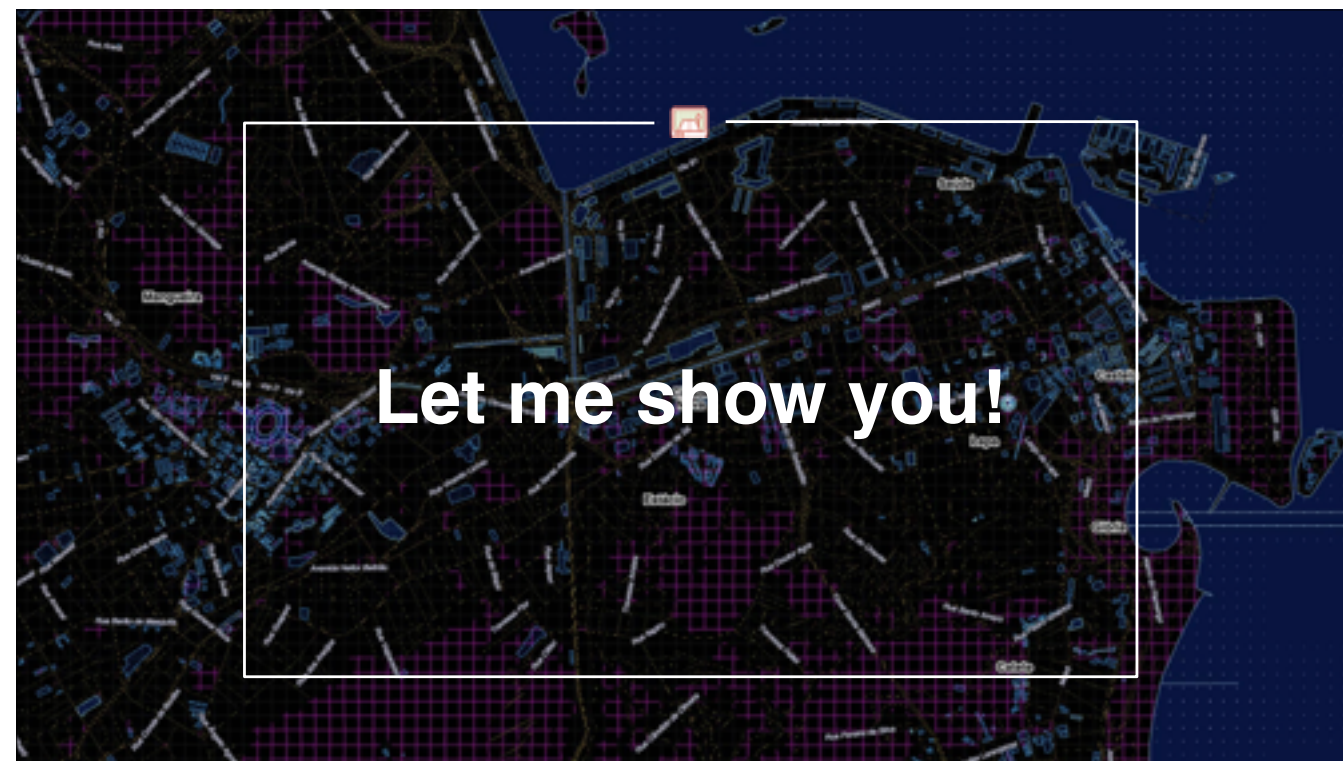
whosonfirst.mapzen.com

- open gazetteer, which is a collection of places
- brings together various open datasets
- generates concordances between the datasets
- provides stable integer ids

Mapzen Search



- When we put together Pelias and open data, we get Mapzen Search
- hosted instance of Pelias with the power of the entire world's open data
- makes our solutions accessible to those that don't have resources to host their own instance



Mapzen Search product page
documentation page
Leaflet plugin tutorial

Mapzen Search

<https://mapzen.com/projects/search>

Mapzen Search API Documentation

<https://mapzen.com/documentation/search/>

Add the Mapzen Search geocoder to a map

<https://mapzen.com/documentation/search/add-search-to-a-map/>