



flight



Profile Geo-Enrichment

Diana Abbas

Software Engineer, Twitter | @dianaa

GROVE STREET

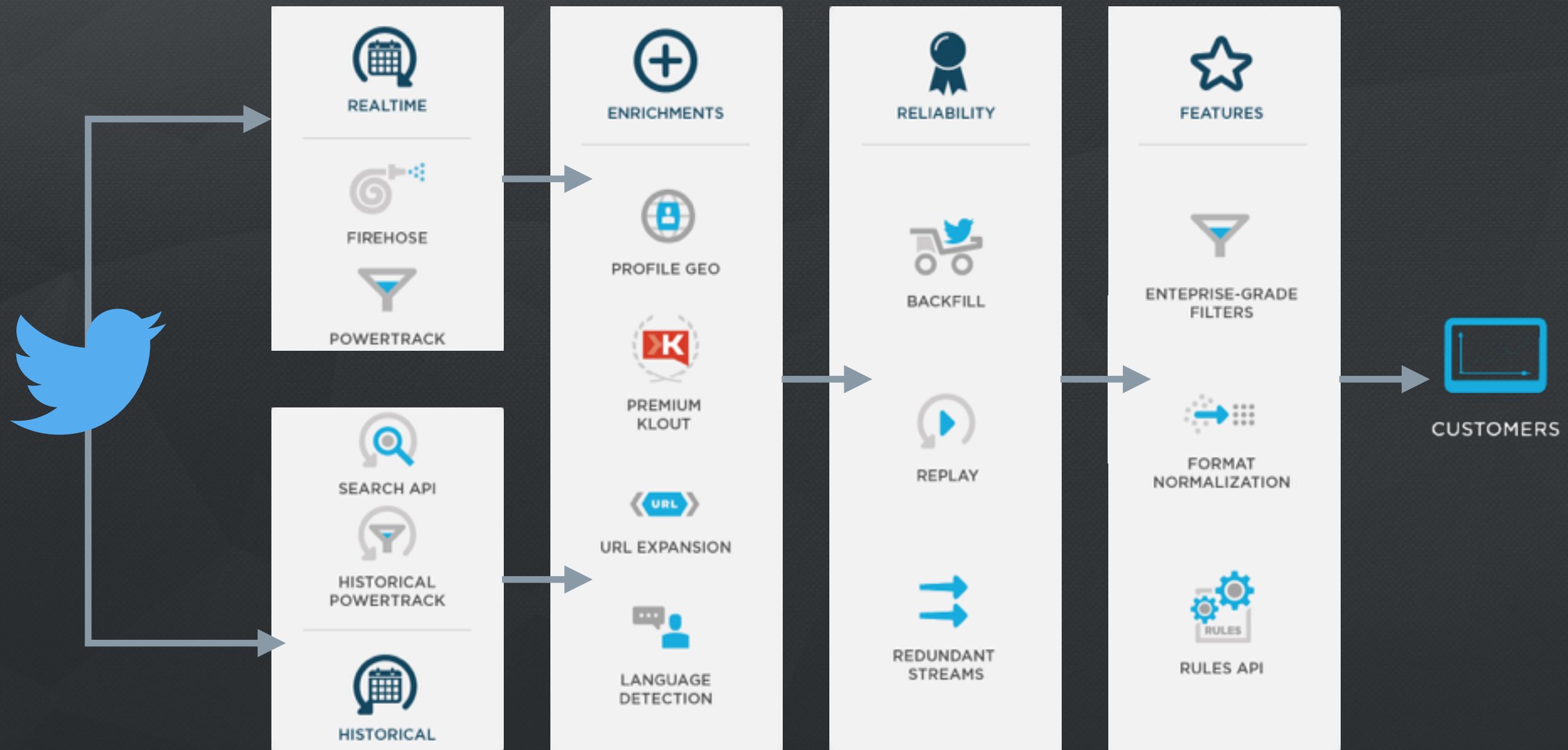
MARKET STREET

MISSION STREET



Why Geo-location?

The Making Of Enterprise Data



Geotag Feature



Verizon 3:55 PM 55%

Remove Tag location Done

Search locations

Boulder, CO ✓

Foundry Group
1050 Walnut St · 150 ft

Twitter
1050 Walnut St #115 · 60 ft


TechStars HQ
1050 Walnut St · 80 ft

Rio Grande
1101 Walnut St · 200 ft

Nor1
1050 Walnut St Ste 202 · 70 ft

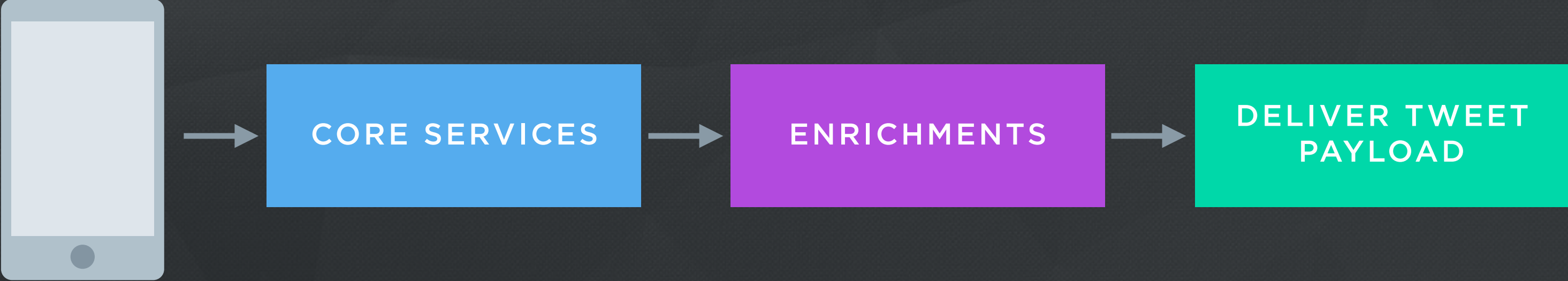
Amante Coffee
1035 Walnut St · 150 ft

Mandelbrot HQ
1050 Walnut St · 100 ft

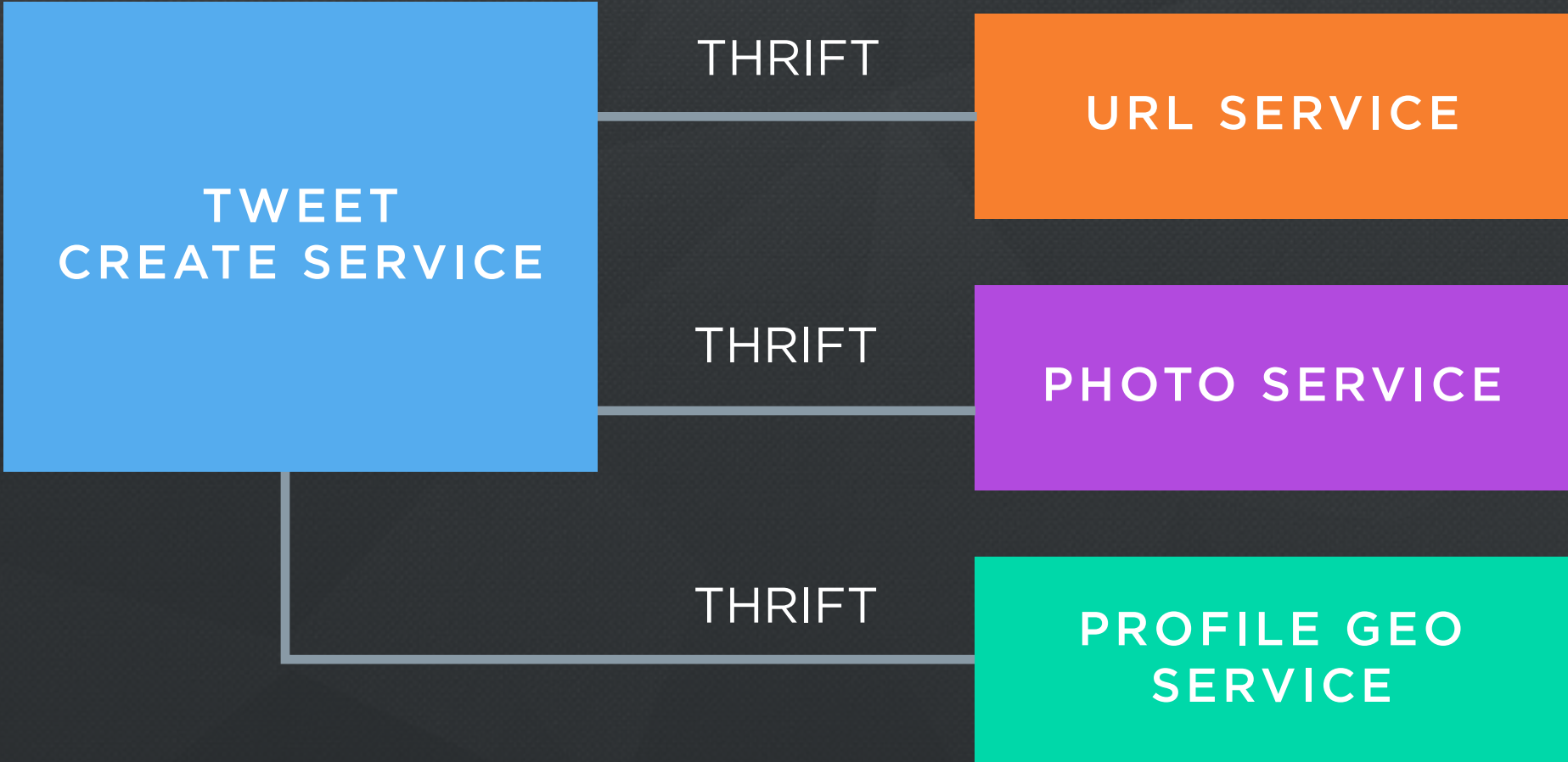


Diana A Boss
@DianaA
aspiring to meet michael and the last unicorn in neverland. never grow up. engineer [@Twitter](#)
Boulder, CO
Joined August 2009

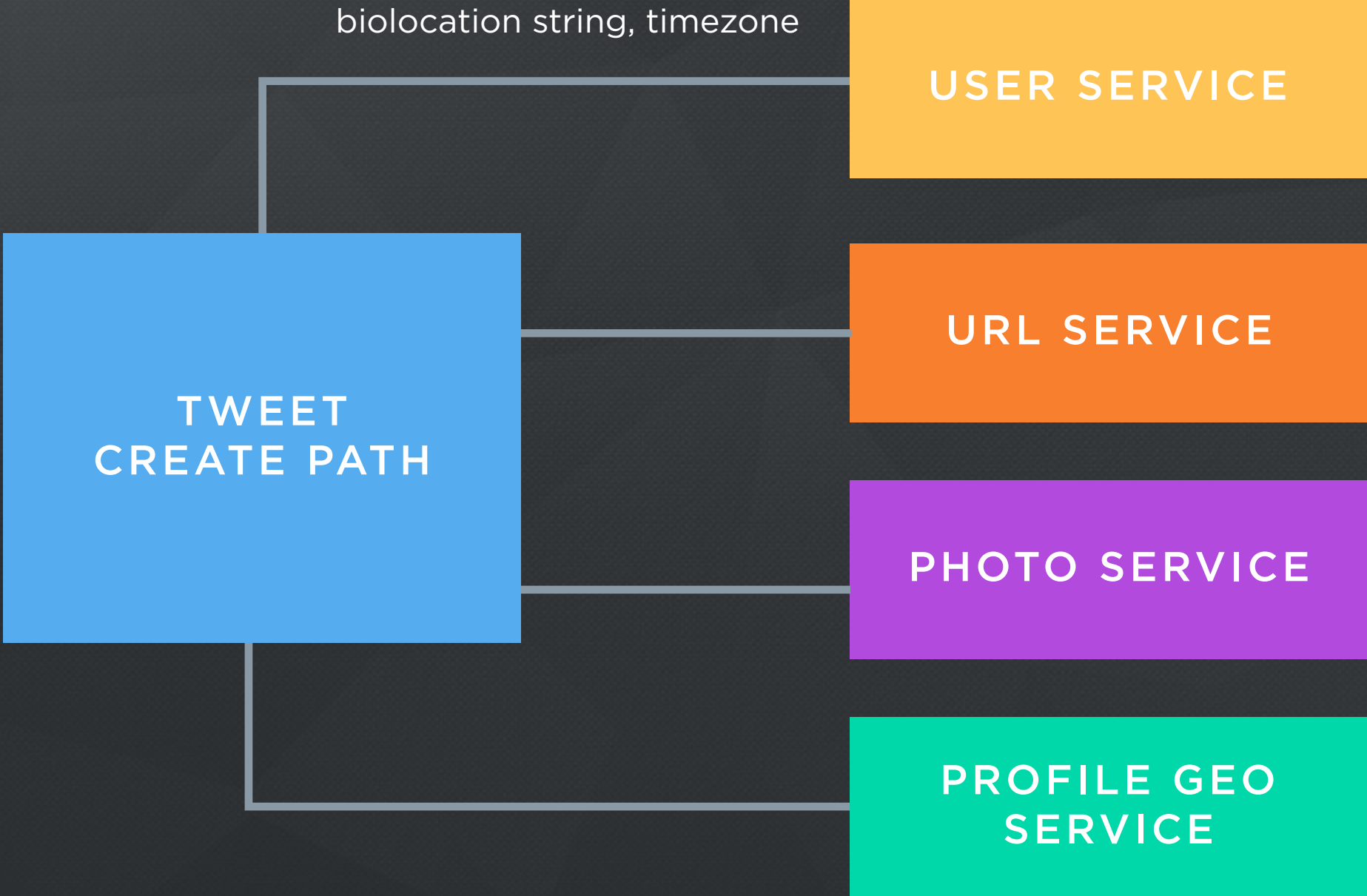
Tweet Delivery Path



Core Services



Core Services



Geonames



Geo Entity

```
struct GeoEntity {
  1: required string entity_id           // unique ID of the entity from the geonames dataset
  2: required string display_name       // something like: Boulder, Colorado, United States
  3: optional GeoPoint geo_point        // center point of the geonames entity
  4: optional string timezone           // time zone name. string from the TimeZone java util
  5: optional string country_code       // ISO value representing the country
  6: optional string region              // usually state or province (US and Mexico)
  7: optional string sub_region          // next largest admin division, adm2 from geonames
  8: optional string locality           // incorporated city or town political entity
  9: optional i32 population
}
```

Profile Geo Candidate

```
struct ProfileGeoCandidate {  
  1: string geoEntityId      // refers to a GeoEntity  
  2: optional GenesisTweet genesisTweet // tweet containing bio location string  
  3: optional LocationMetaData metaData
```

String Normalization

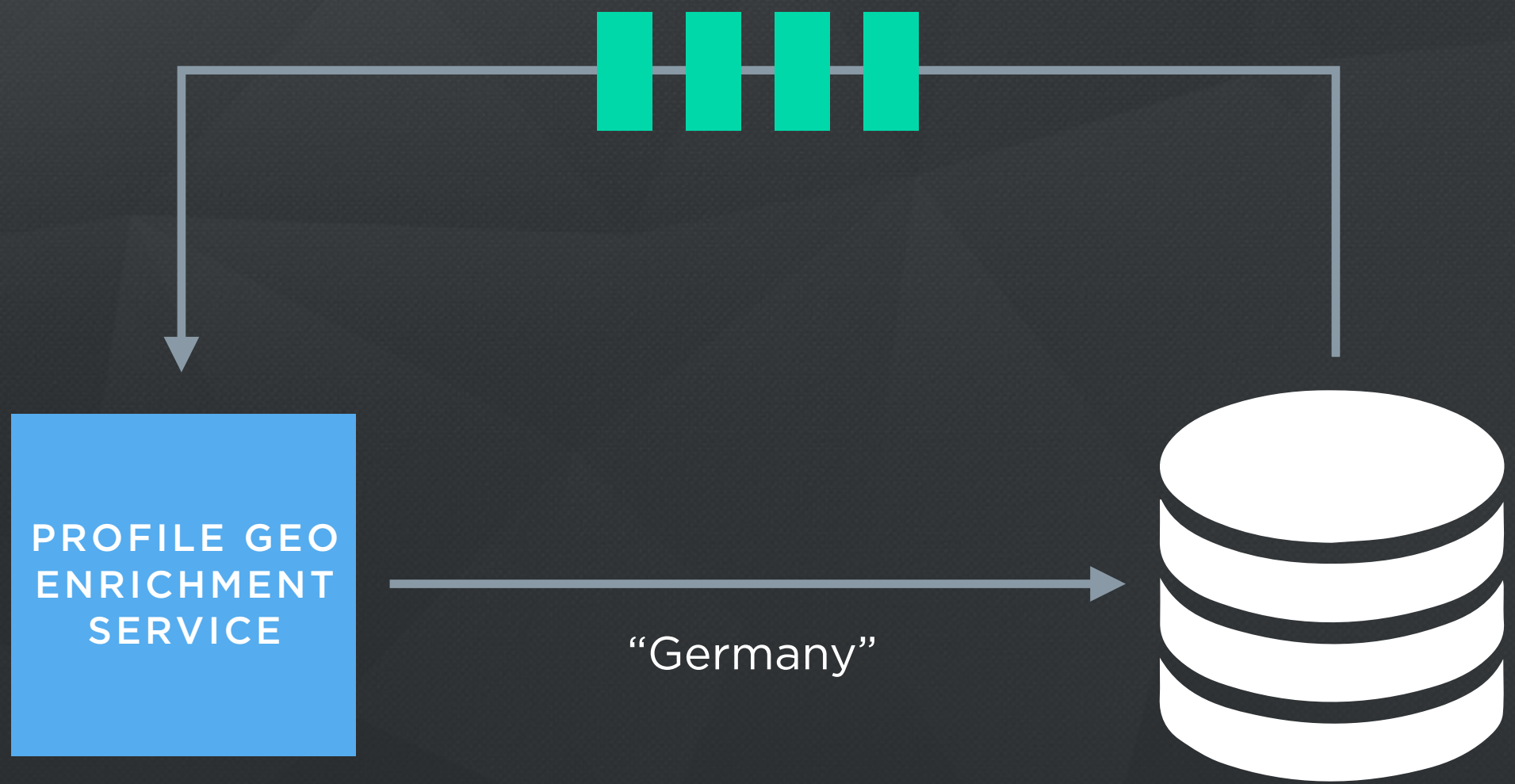
“Republic of
Germany”

“Deutschland”

“Germany”



Profile Geo Enrichment Service



Selection Heuristics



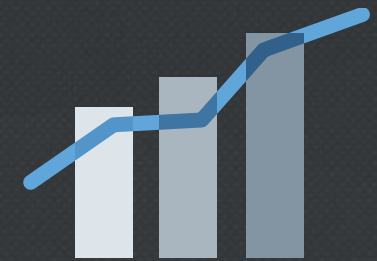
Candidate closest to
lat long of request
within 100 km



Matches exactly on
the country name



Matches on a major
city, population
> 200,000



Highest relevance
score > 80



Highest good to bad
user metadata ratio



Best country match

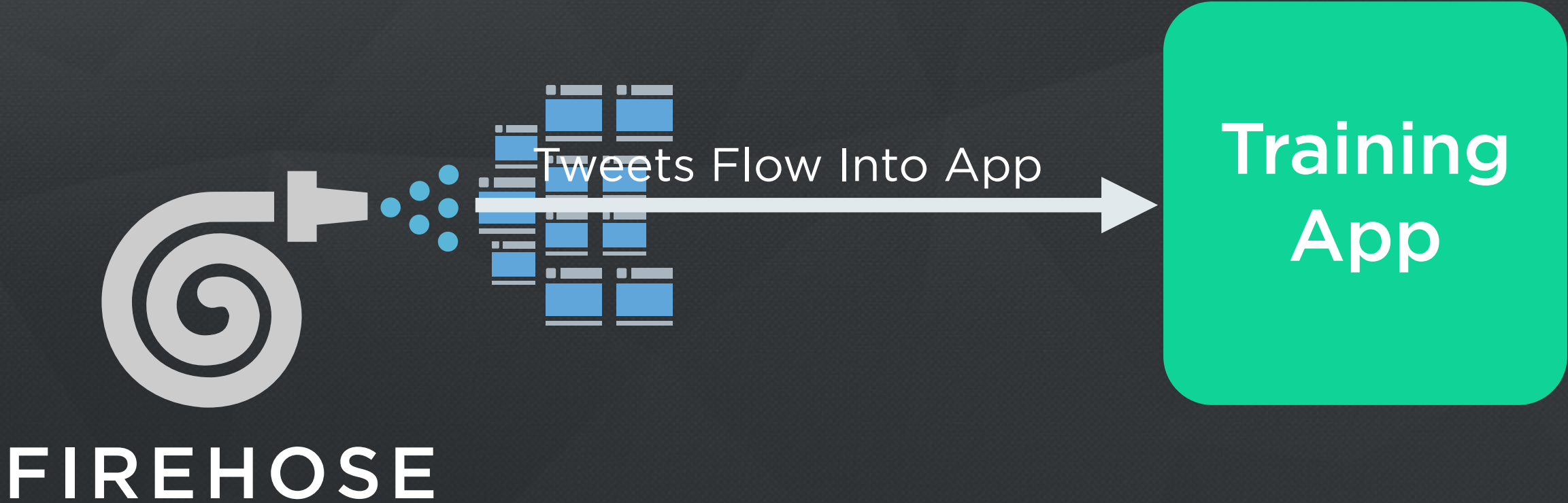


Best time
zone match



Highest relevance

Training Model



Lat Long Matching

FIREHOSE



PROFILE GEO
TRAINING



LUCENE INDEX

Search

Match Popularity

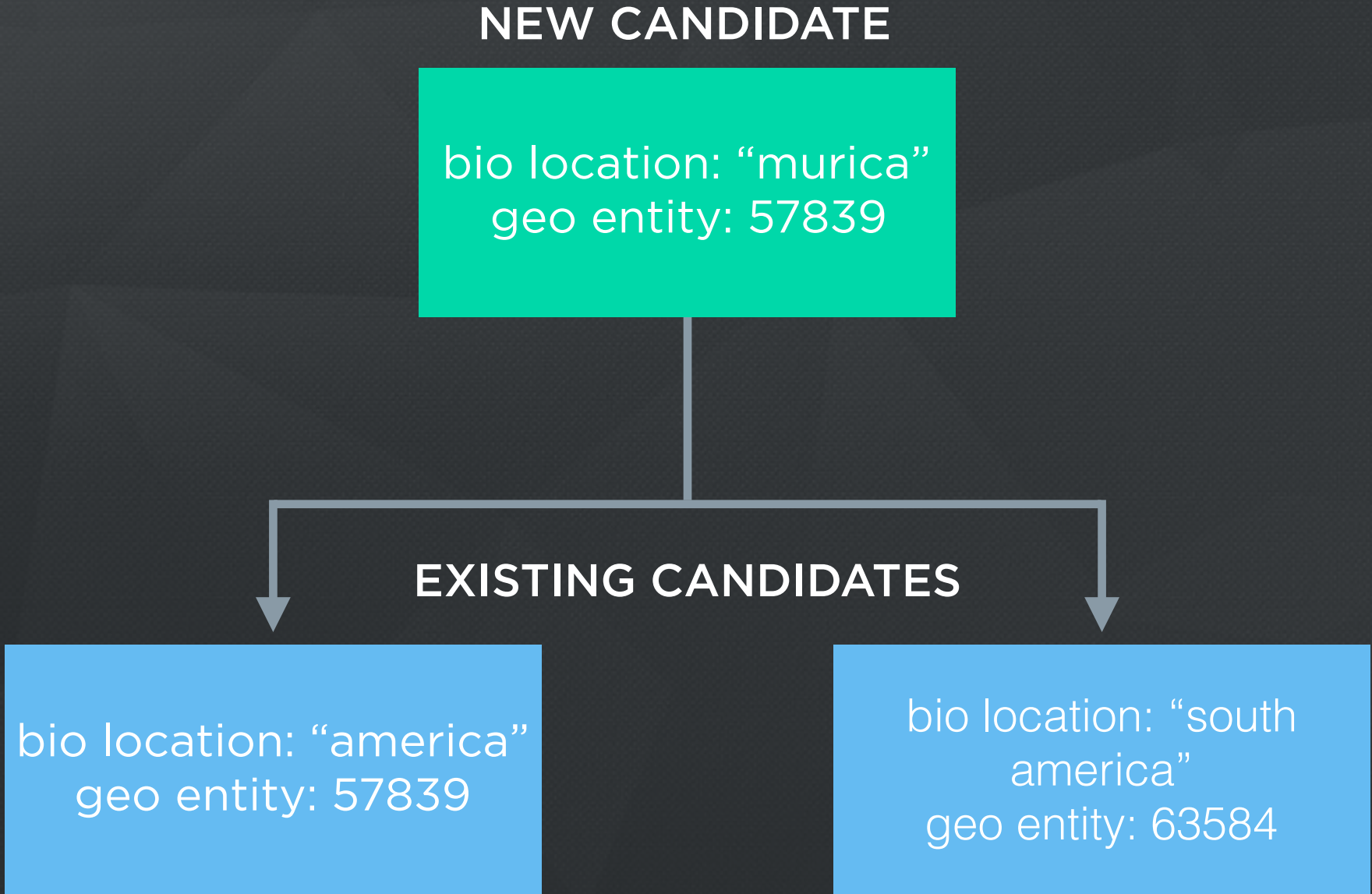
KEEP TRACK OF USERS
TWEETING WITH THE
SAME BIO LOCATION

COUNT GOOD MATCHES

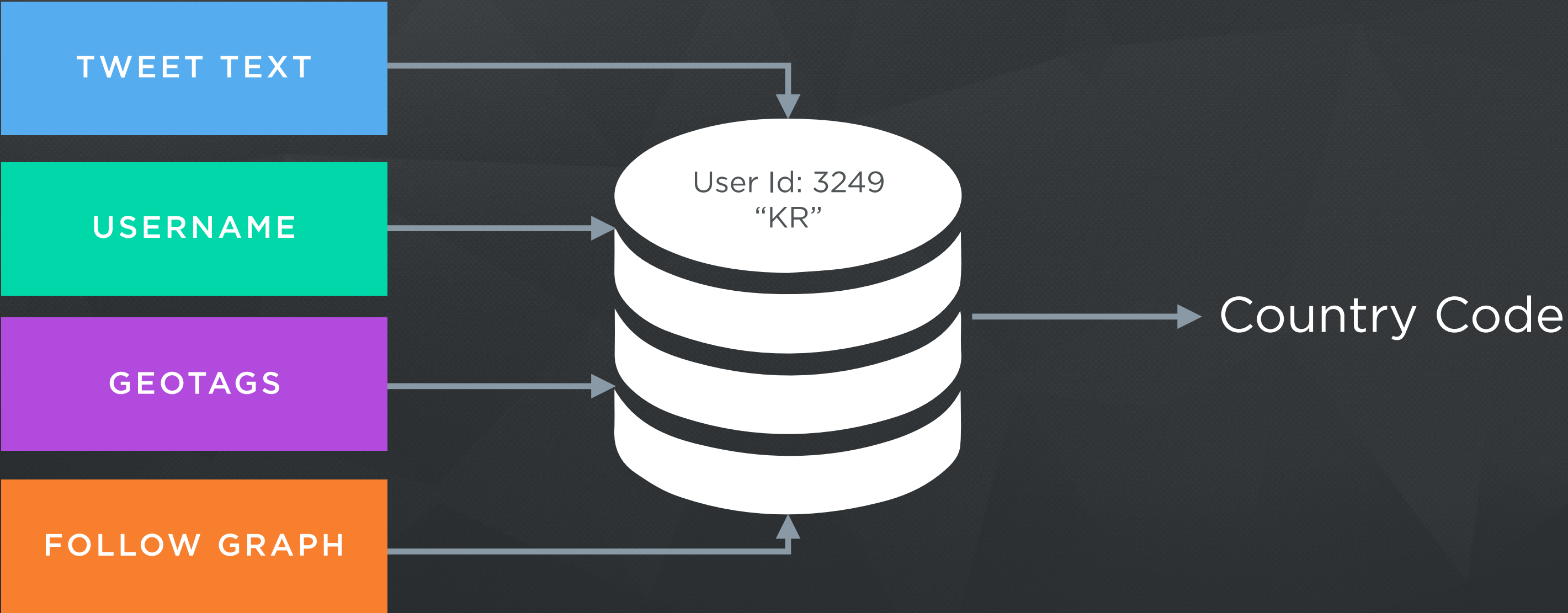
COUNT BAD MATCHES

FILTER WITH RATIO
METADATA

Update Good And Bad Matches

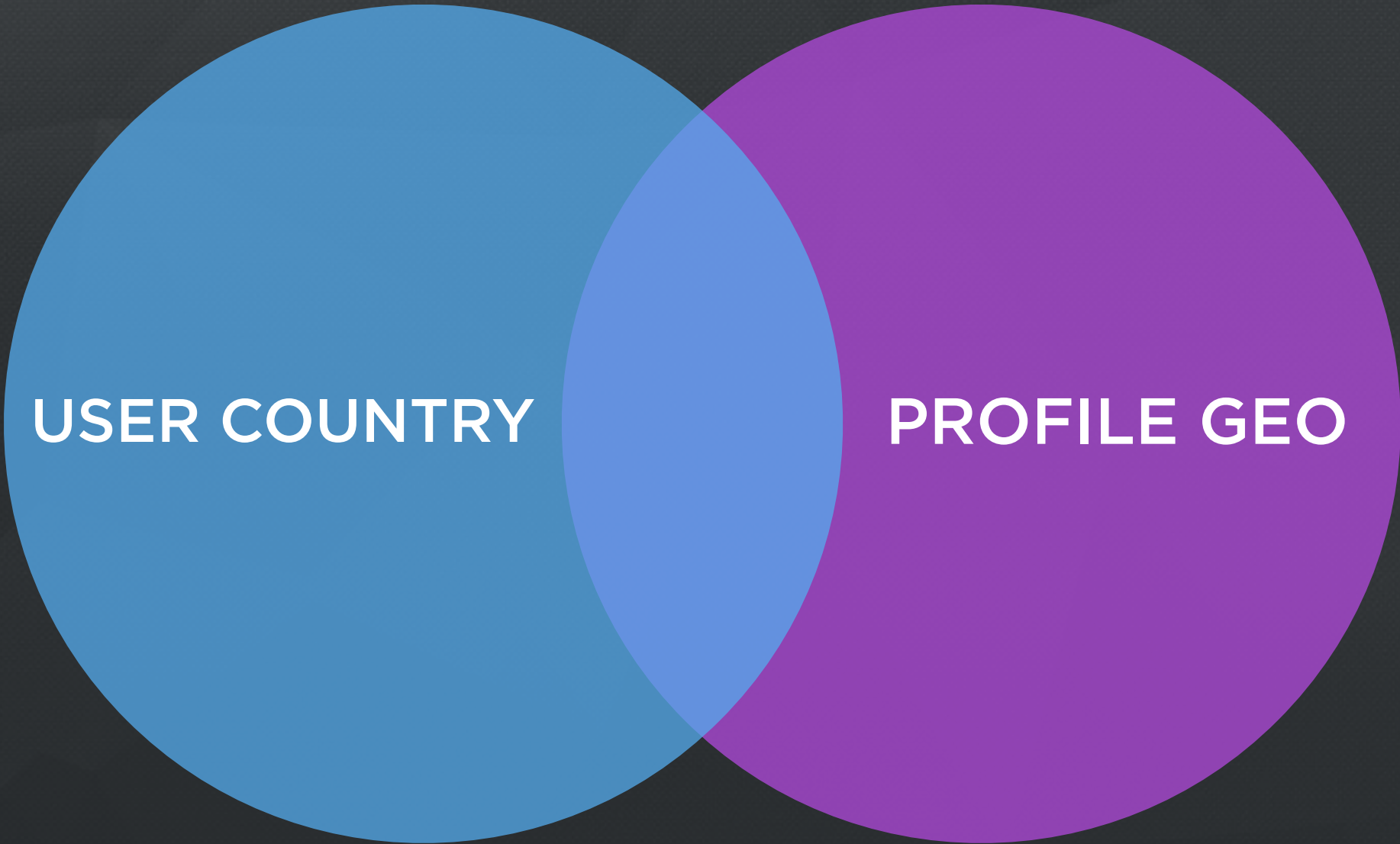


User Country Model



Candidate Filtering

“Birmingham, AL vs Birmingham, UK”

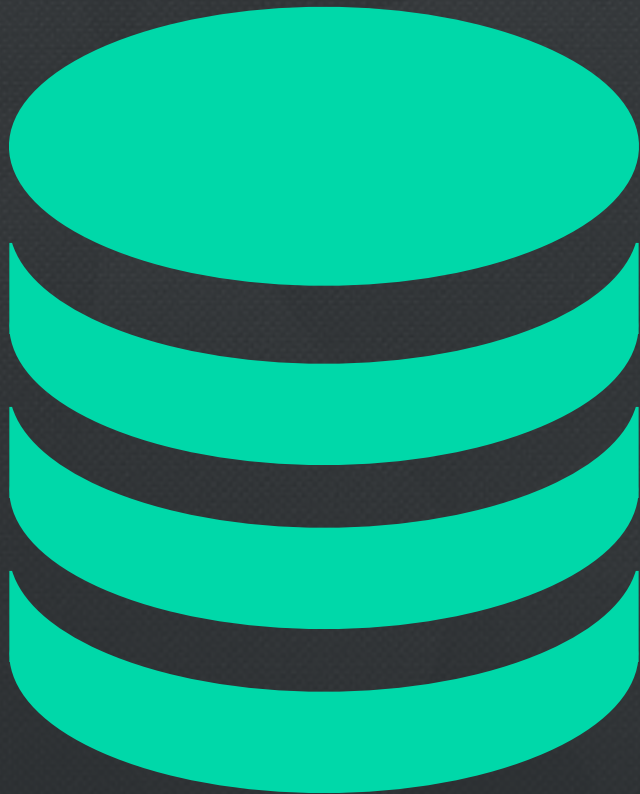


Read Path

Full archive
search

Historical

Model Evolution

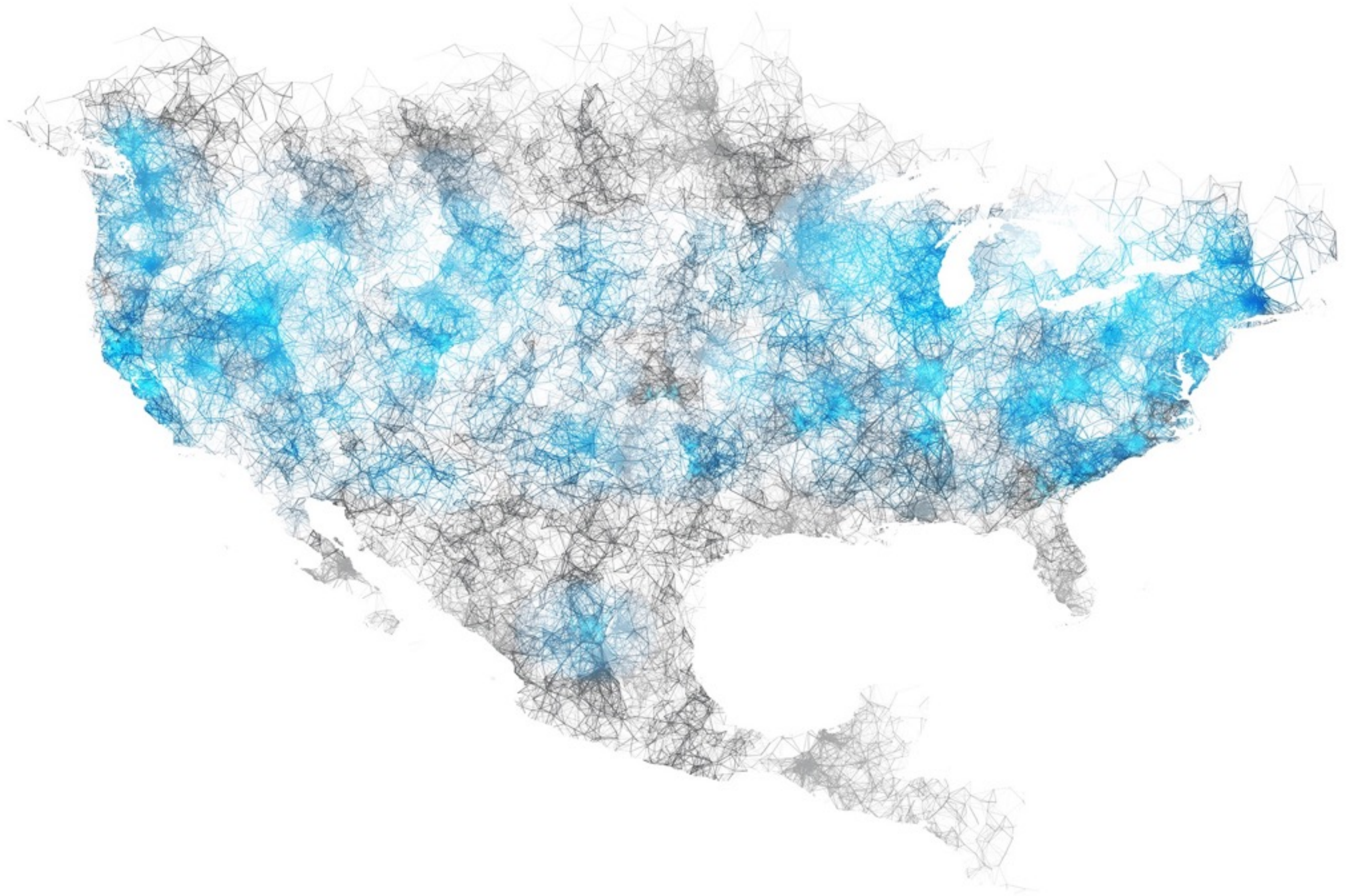


LESSONS LEARNED

Location Prediction is Not Easy

Hard to Test Accuracy

Many Use Cases





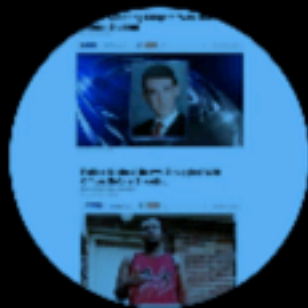
#Ferguson

Mentions on Twitter of #Ferguson

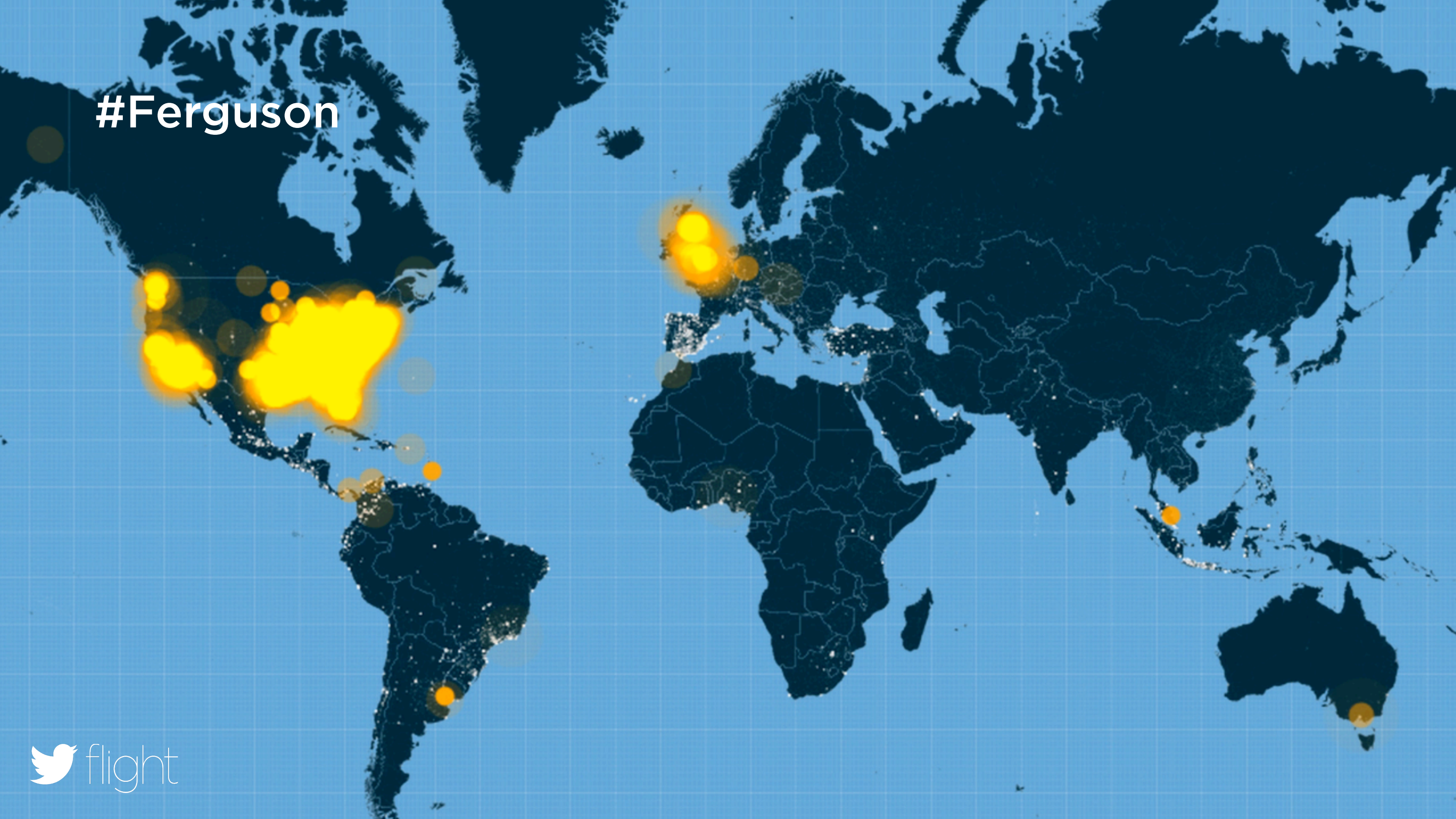
by Elana Zak @elanazak

5

Total Tweets
Aug 9, '14 - Aug 9, '14
GMT+4



#Ferguson



Thank You

@dianaa