Collecting The World's Open Address Data

Ian Dees

https://openaddresses.io
@iandees @openaddr

The problems we're trying to solve

- Addresses are the human interface to geographic data
- A map isn't usable by humans until they can locate themselves
- Street address is where everyone starts
- Geocoders are the software that translates the address to somewhere on the map

The problems we're trying to solve

- Lack of geocoders
  - Google's is a good example
  - Decent API, above average results
  - Opaque pricing, license restrictions
  - Mapzen Pelias, Mapbox, Nominatim
- Lack of data
  - Improve access to data
  - Promote open data
- Sharing data improvements

Existing projects

OpenStreetMap

- Mature dataset, but “deep” map data like addresses aren’t as easy to add
- ODbL license leaves some questions about legal expectations of geocoding results
- 61M addresses, 1.9M ranges worldwide
- 30K monthly active mappers

Existing projects

OpenAddresses.org

- Very similar, but has gone quiet
- Focus was on European countries
- No up to date output

Existing projects

OpenAddressesUK

- Collected addresses in the UK in reaction to the national address set becoming a public-private partnership
- Was not collecting geographic data
- Has gone dormant

Our Project

- Started as a Google Spreadsheet of potential OSM import sources
- Not enough time to shepherd the imports

Our Project

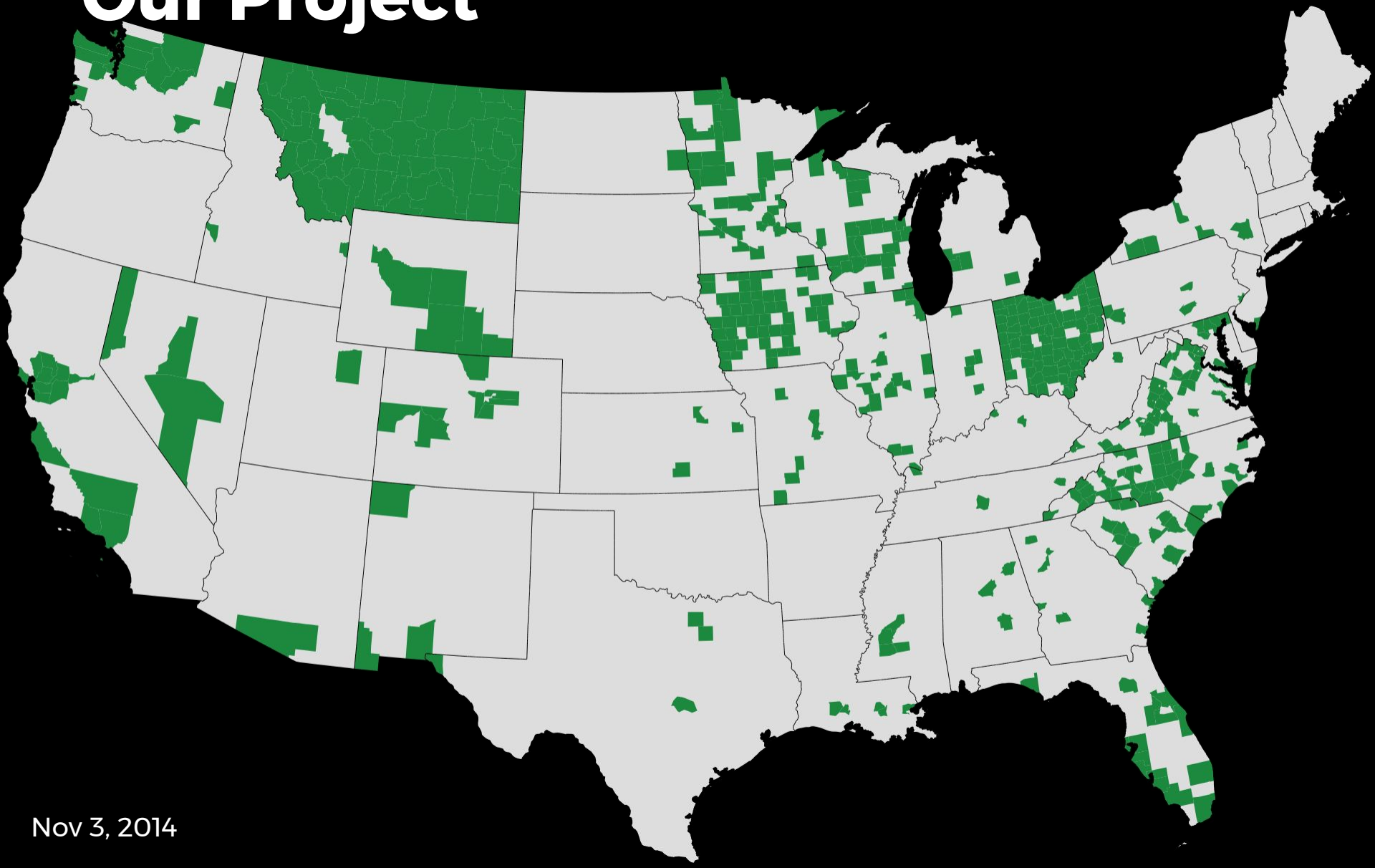
- A GitHub project that lists ~2000 URLs pointing to open data of various sorts
- Software to download, cache, and collate a unified output from that collection of local, authoritative data sources

Our Project

https://openaddresses.io

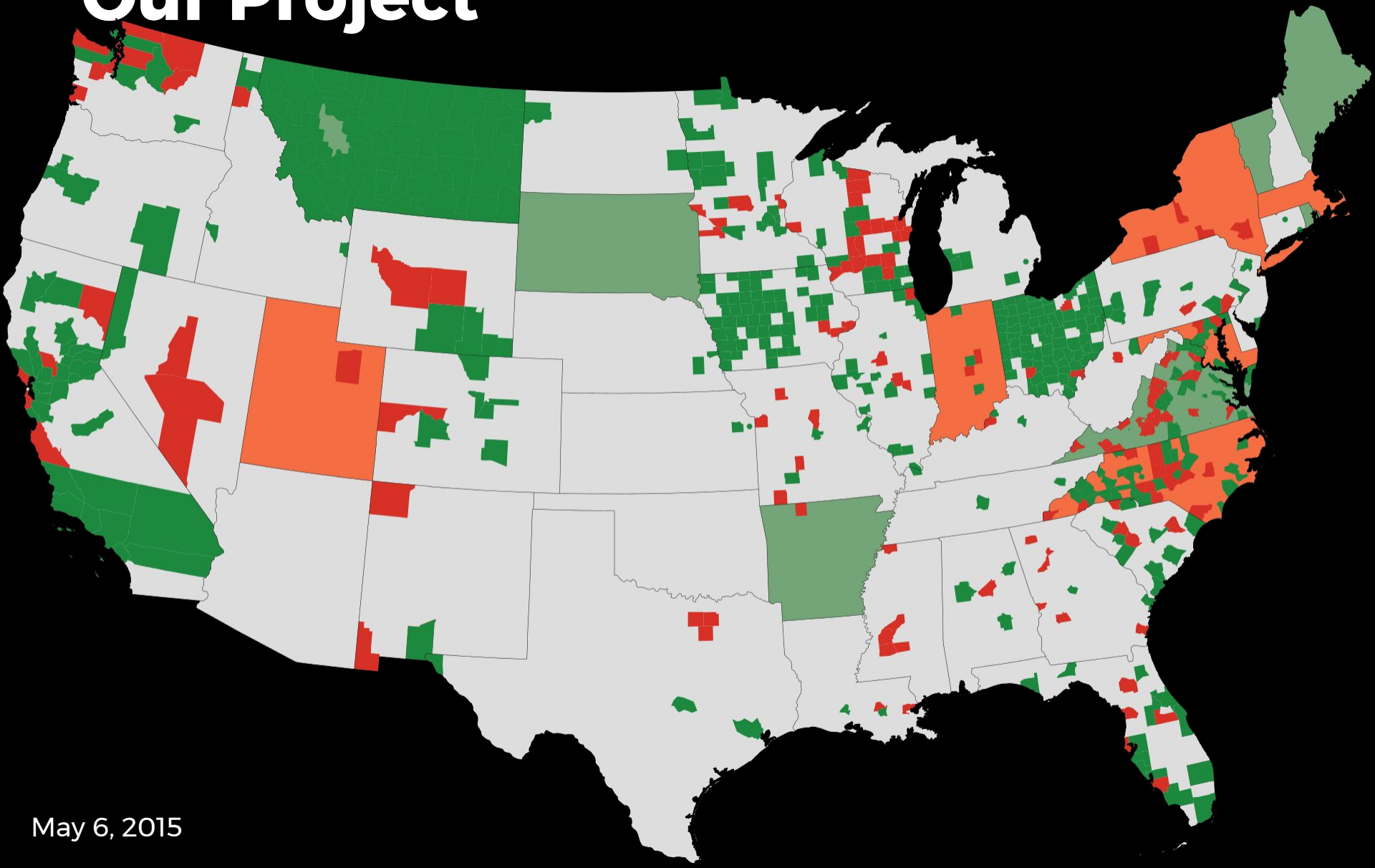
- Currently includes the bare minimum: lat, lon, house #, unit, street, city, state, postcode
- Growing fast, lots of room for improvement
- 250M 472M points, 64 113+ contributors

Our Project



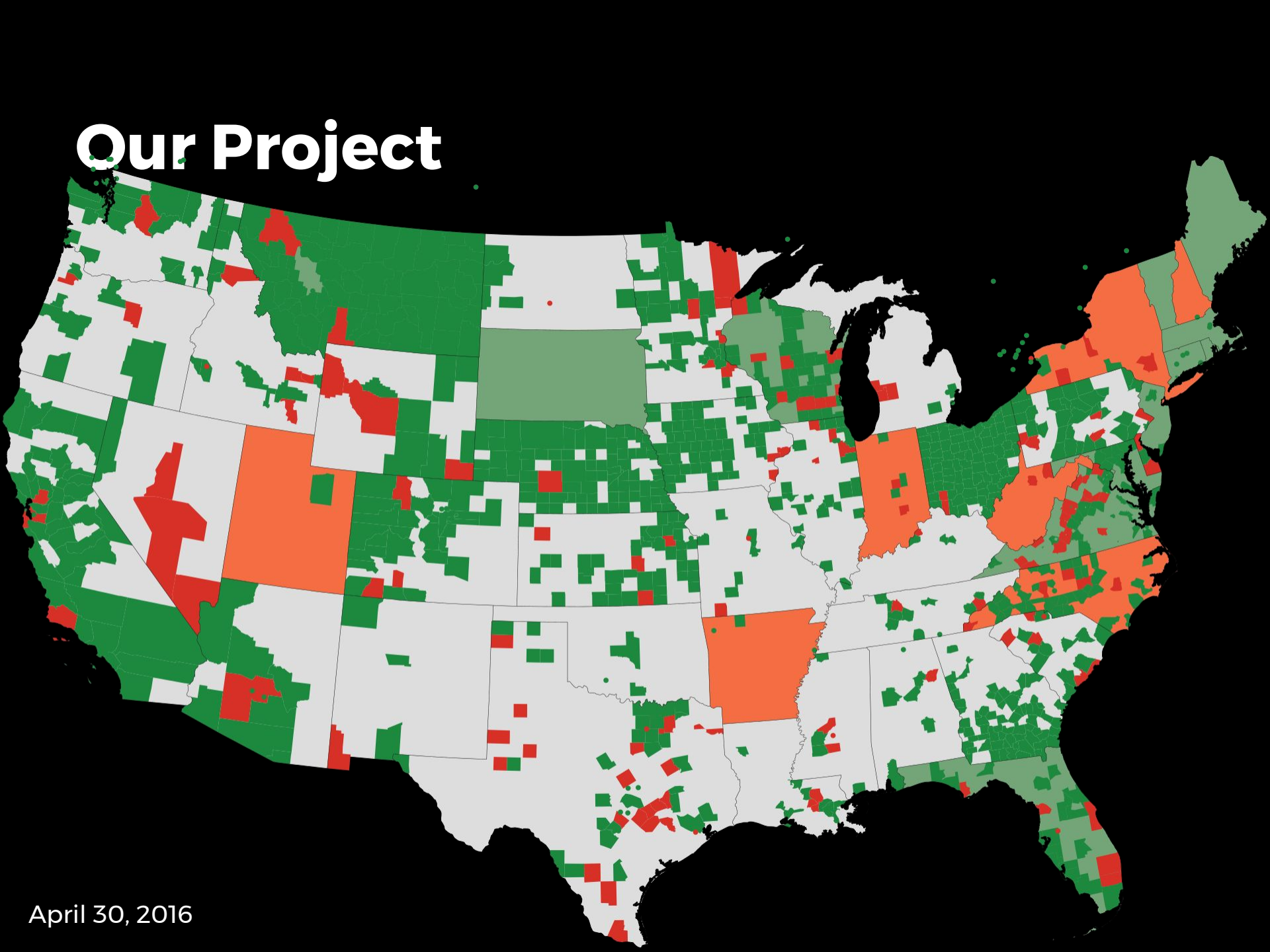
Nov 3, 2014

Our Project



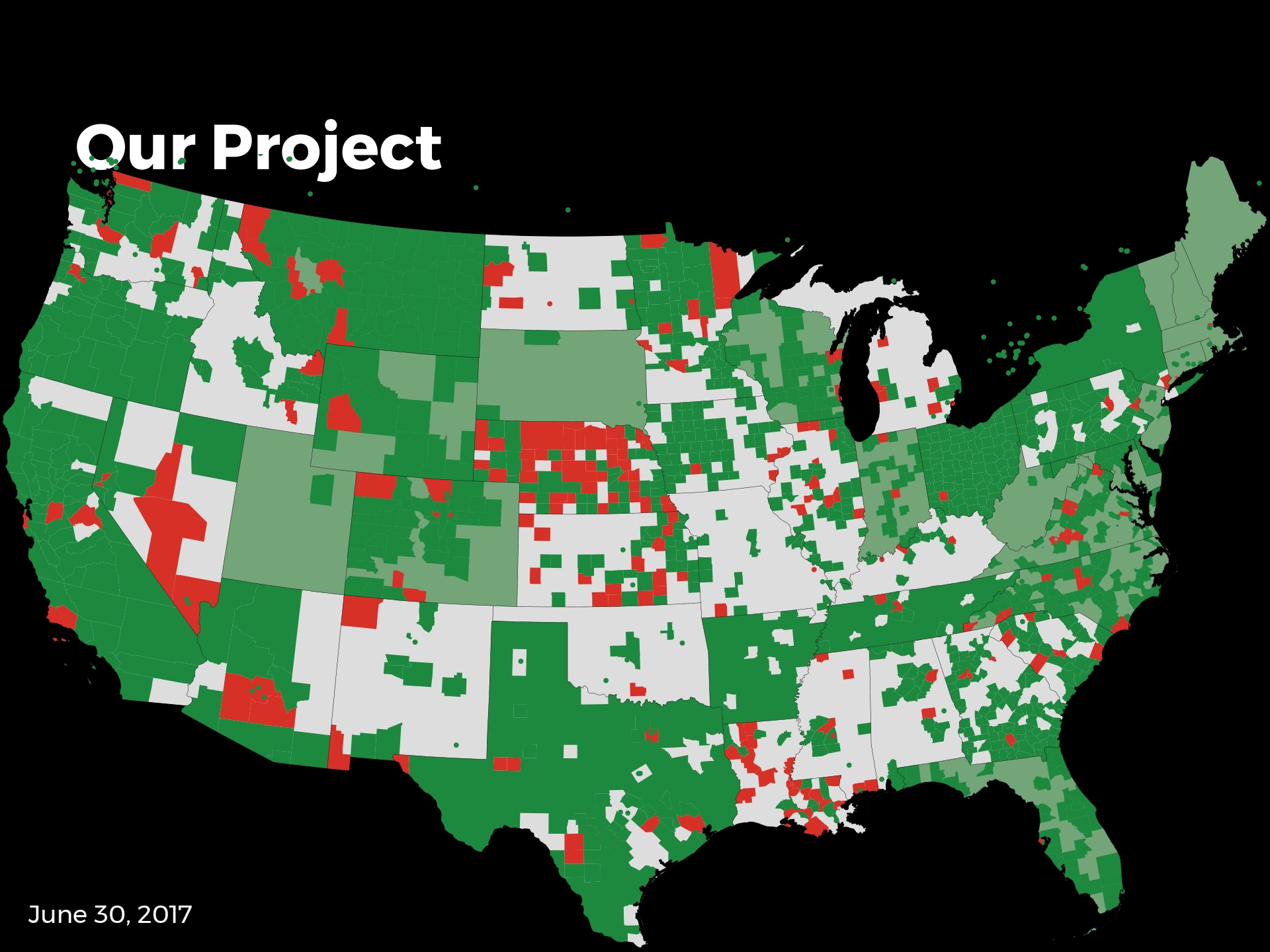
May 6, 2015

Our Project



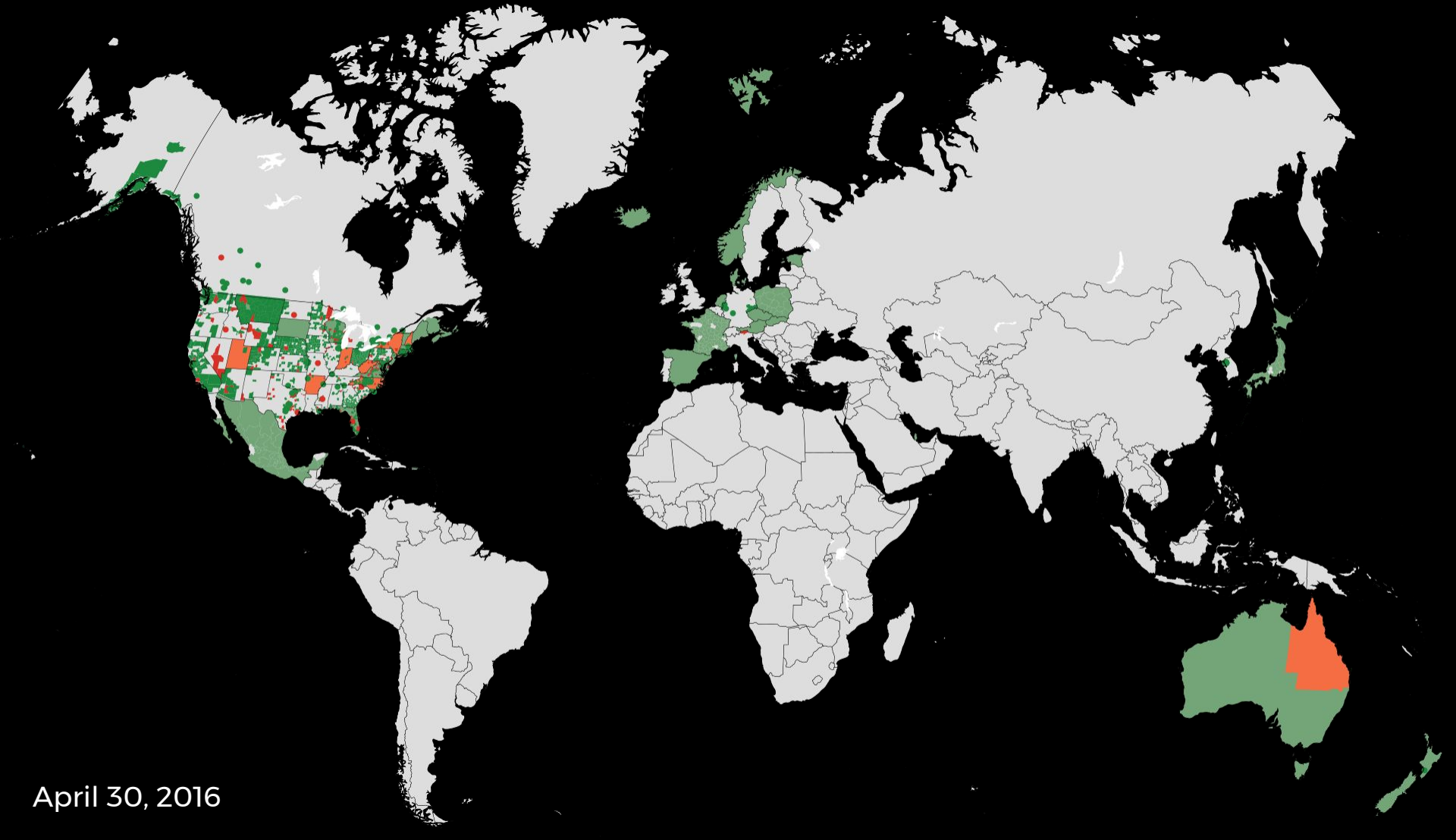
April 30, 2016

Our Project



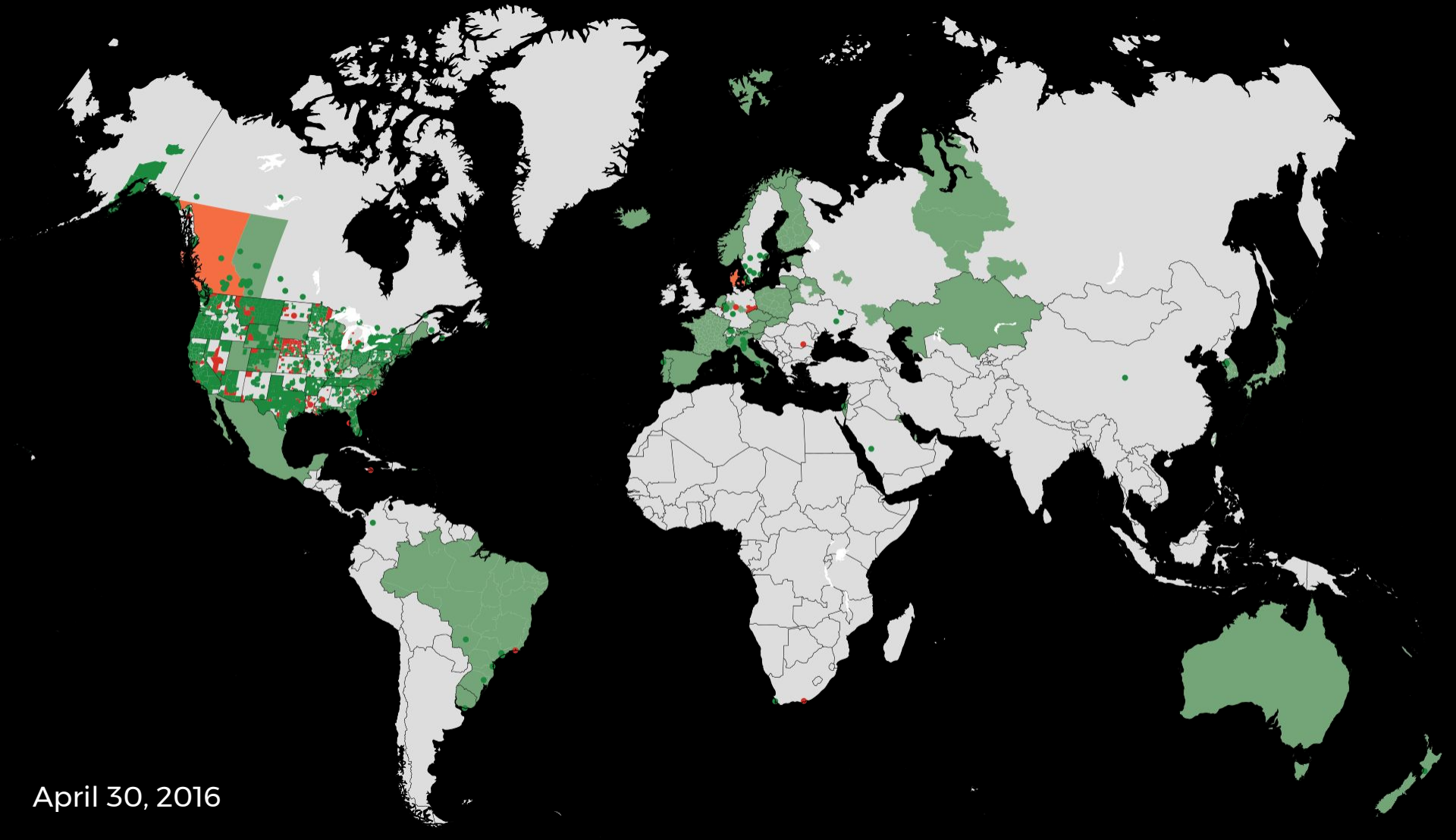
June 30, 2017

Our Project



April 30, 2016

Our Project



April 30, 2016

GitHub-based Infrastructure

- All contributions are made via Pull Requests
- Allows conversation and community building
- Automated checks are run to give feedback
- Example of pull request workflow:
  - https://github.com/openaddresses/openaddresses/issues/344
  - Add a new source
- Please steal this idea!

Simple Python Backend

- Each source file has a data URL with instructions for how to download
- Supports HTTP, FTP, ESRI downloads
- Converts geodata from the source into a CSV
- Mapping between source and output column names
- Merge of individual sources into an output ZIP

Our Answer To The Schema Problem

- Lots of existing address standards
- Our simple, minimal output is immediately useful to current consumers
- Caching full-fidelity source data means we can add to our output at any time
- “unit” field was added and sources were updated over time to improve output

Licensing

- Most sources consider their data public domain or not copyrighted
- Some require attribution, a few require share-alike
- There are flags in sources to indicate these requirements
- We group outputs into share-alike and non-share-alike

Example Usage

- Mapzen's Pelias project is a prime example of using OpenAddresses
- Quick Demo

Where to next?

- Crowd-sourcing points isn't sustainable
- Citizens are already paying for the collection/creation
- Work with local government to upload their own data
- Use OpenAddresses as an example for those looking to create open data
- Lots to do internationally

Where to next?

Feedback mechanism

- OpenAddresses.io could accept information from data consumers/geocoders and give it back to data sources

Where to next?

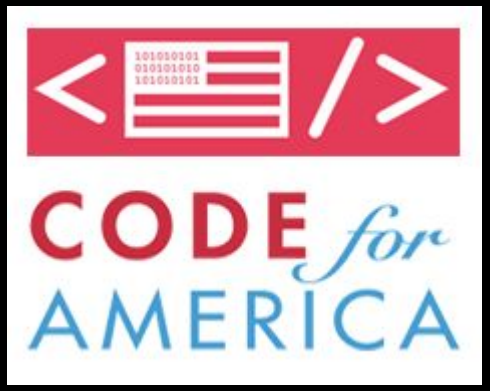
Licensing

- We leave it up to the consumer to understand licensing implications
- Not sustainable, makes it harder to use data
- We should do a better job tracking a source's licensing info

How to help

- Find your local data
- Tell us about it (issues or pull requests)
- Tell your locality about us
- Download the data and play with it

Special Thanks



Hosting and processing is powered by AWS with help from Code for America

U.S. Open Data funded a data bounty
https://github.com/openaddresses/openaddresses/issues/265

Any Questions?

Electronic Mail:
hello@openaddresses.io

Electronic Bird Noises:
@iandees