Data Science for Smart Cities

CE88

Prof: Alexei Pozdnukhov

Today

Urban data collection, handling and processing.

Agenda:

Lecture 5. Data acquisition: measurement and crowd-sensing.

Lecture 6. Community surveys, population census,
open government data, APIs.

Break

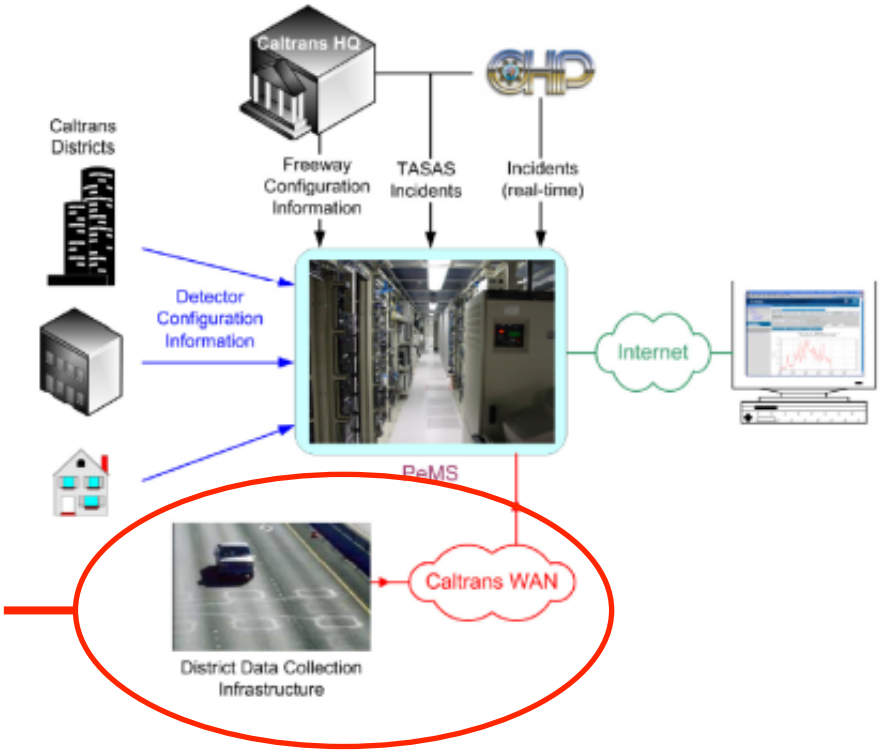
Mini-lab

Data acquisition

Data acquisition is the process of sampling signals that measure real world physical conditions and converting the resulting samples into digital numeric values. Primary way of collecting data on performance of infrastructures.

- performed by a dedicated hardware installation
- may require significant investments for installation and maintenance(!)
- purposefully designed, therefore often includes data quality control
- can be combined with controls (SCADA systems)

PeMS example:



Traffic loops detectors data acquisition

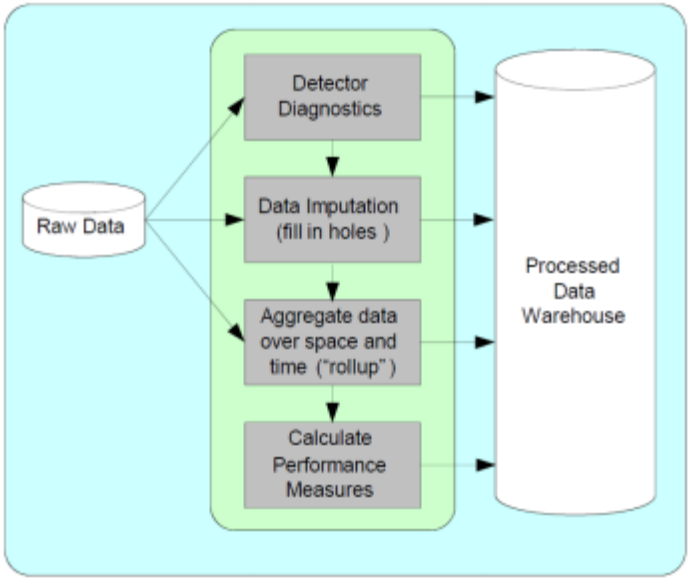
Data quality and data aggregation

Individual sensors occasionally malfunction, stop working or cease sending data. These errors occur for different reasons, and result in gaps in data. Well designed data acquisition systems include data quality diagnostic tests and detector 'health' information.

Common strategies to deal with the missing data situation are:

- insert a default value indicating that the reading is not available
- impute a 'reasonable', estimated value instead of a real reading

PeMS example



Discussion:

Communication costs vs Computation costs

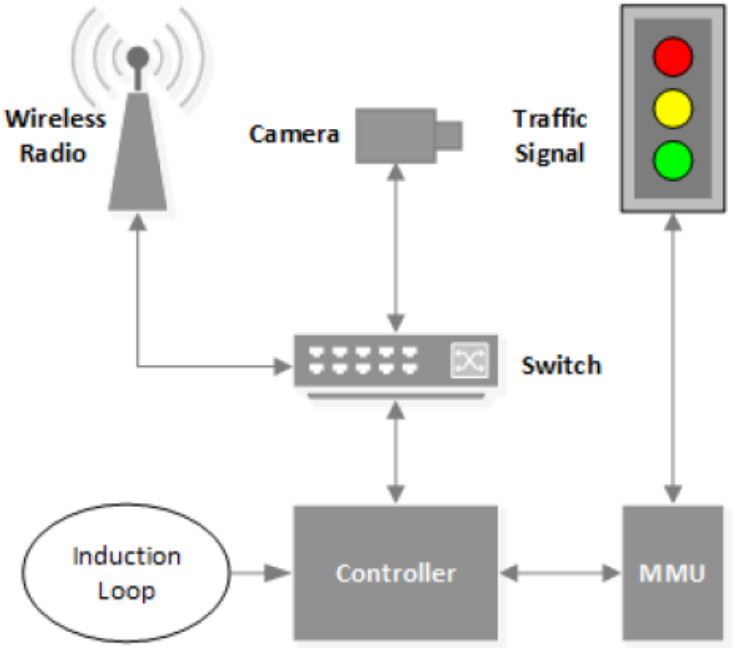
SCADA

SCADA (**supervisory control and data acquisition**) is a type of system for remote monitoring and control of physical infrastructures.

- operates with coded signals over communication channels;
- may be combined with a data acquisition system by adding the use of coded signals over the same communication channels;
- acquires information about the status of the remote equipment for display or for recording functions.

Example

http://thehackernews.com/2014/08/hacking-traffic-lights-is-amazingly_20.html



How do we collect data about people?

US Census Bureau

http://www.census.gov/

US Census Bureau

What they do

The Census Bureau's mission is to serve as the leading source of quality data about the nation's people and economy.

How they do it

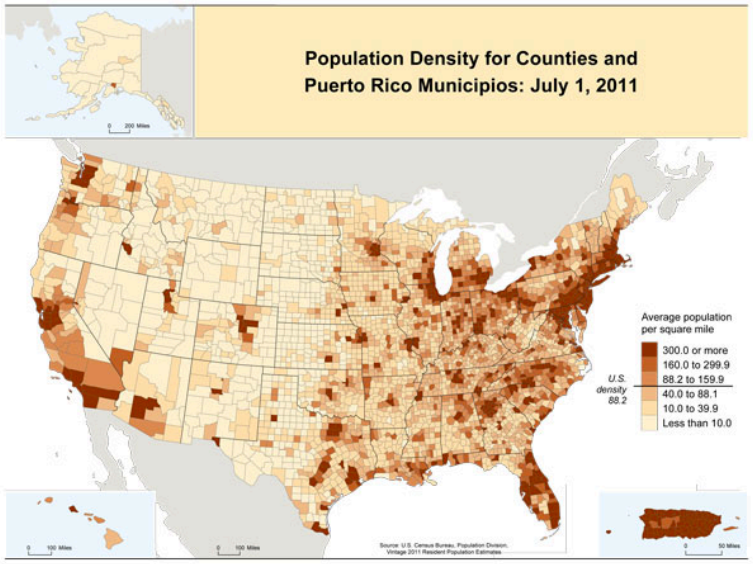
The US Census Bureau conducts more than 130 surveys a year, including:

- Household surveys, the American Community Survey
- Business Survey, Annual Retail Trade surveys
- and many more.

US Census Bureau

Census is used to distribute more than $400 billion in federal funds to local, state and tribal governments each year. Census data informs how states and communities allocate funding for:

- Neighborhood improvements
- Public health
- Education
- Transportation
- Much more



For example: to make planning decisions about community services, such as where to:

- Provide services for the elderly
- Build new roads and schools
- Locate job training centers

and also:

- To determine the distribution of Congressional seats to states.
- Used to apportion seats in the U.S. House of Representatives
- Used to define legislature districts, school district assignment areas and other important functional areas of government

American Community Survey

http://www.census.gov/programs-surveys/acs/

U.S. Department of Commerce | Blogs | Index A-Z | Glossary | FAQs

United States Census Bureau

Search

Topics Population, Economy | Geography Maps, Products | Library Infographics, Publications | Data Tools, Developers | Surveys/Programs Respond, Survey Data | Newsroom News, Blogs | About Us Our Research

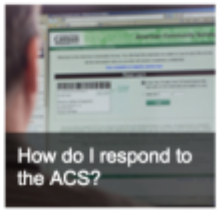
American Community Survey (ACS)

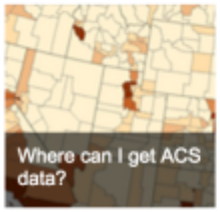
- About the Survey
- Respond to the Survey
- News & Updates
- Data
- Guidance for Data Users
- Geography & ACS
- Technical Documentation
- Methodology
- Library
- Operations and Administration
- Contact Us

Your Help Needed

Participate in a brief online usability test to improve this website

The American Community Survey helps local officials, community leaders and businesses understand the changes taking place in their communities. It is the premier source for detailed information about the American people and workforce.





Latest

Data | News | Events | Library

2014 Data Release

Learn more about the latest ACS 5-year data release on December 3rd, including changes for this release and links for more information.

2014 State Ranking Tables

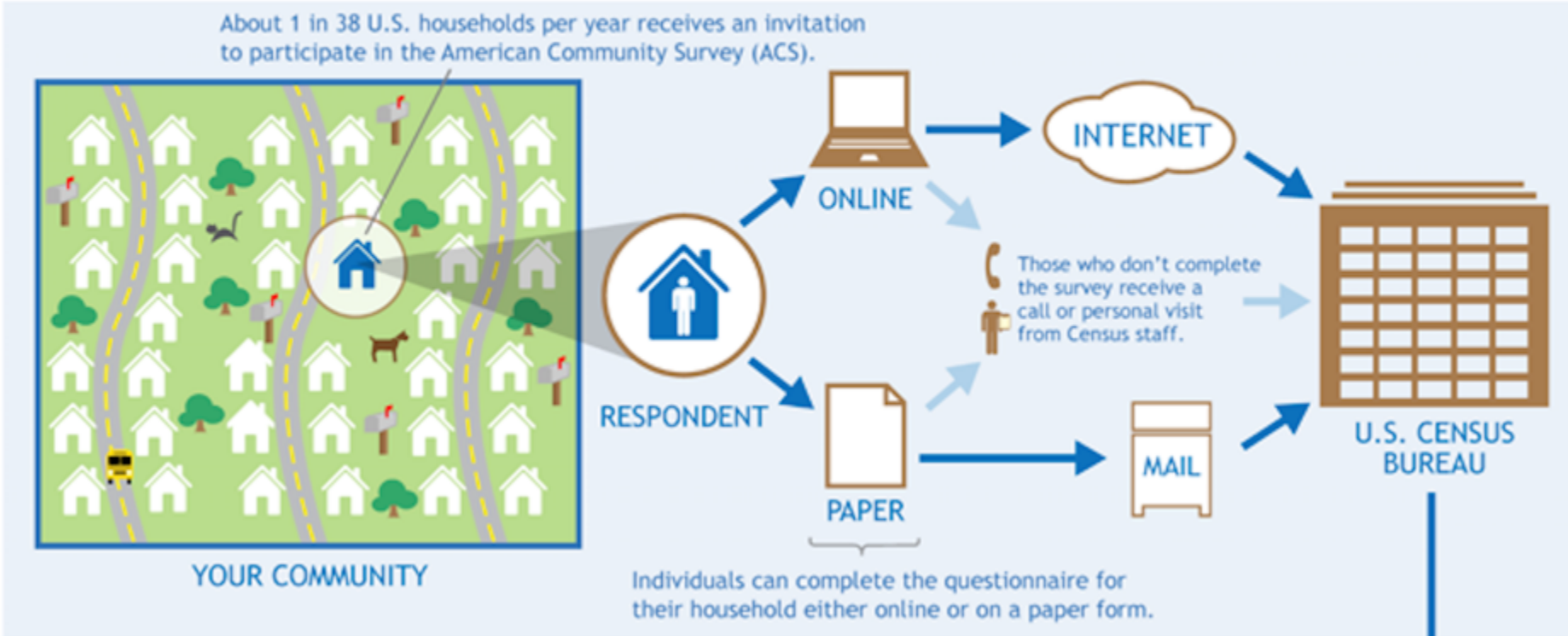
Get 2014 ranking tables that order the nation, 50 states, DC, and Puerto Rico, based on 89 different demographic, social, economic, and housing measures.

Data Tables and Tools

American Community Survey

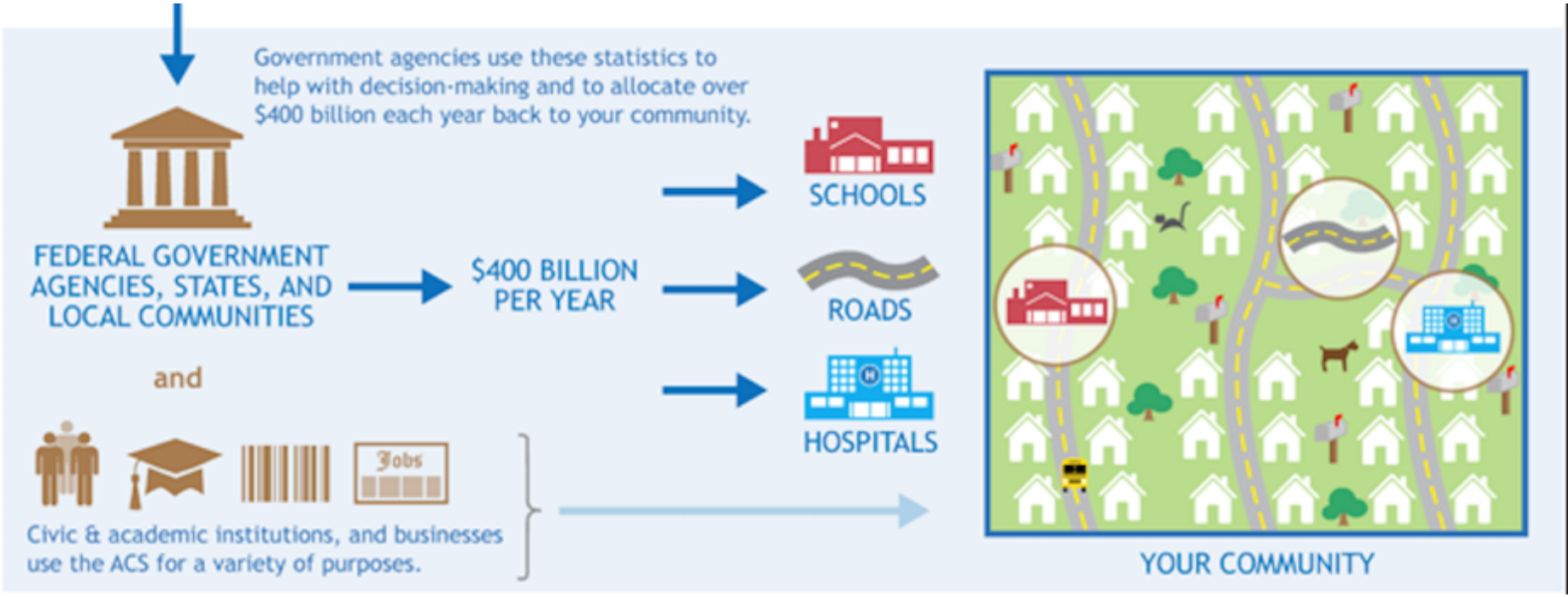
The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about the nation and its people. Information from the survey generates data that help determine how more than $400 billion in federal and state funds are distributed each year.

How the American Community Survey Works for Your Community



American Community Survey

American Community Survey



National Household Travel Survey

http://nhts.ornl.gov/index.shtml

NHTS National Household Travel Survey
Our Nation's Travel

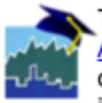
U.S. Department of Transportation
Federal Highway Administration

Home | Analysis Tools | Documentation | Publications | Downloads | Contact Us

About the NHTS

- Introduction
- FAQ
- NHTS Academy
- User's Guides
- Summary of Travel Trends
- Compendium of Uses
- Add-on Program

The NHTS Academy



The NHTS Academy is a series of short, informational videos on NHTS topics of interest to new users and experienced analysts. The Academy has grown to five subject areas with nine videos.

NHTS at TRB

The TRB Task Force on Understanding New Directions for the NHTS sponsored a Poster Session and a Workshop at the 2016 TRB Annual Meeting. The posters and presentations are now online.

2015 Compendium Updated

The NHTS Team has released its latest 2015 Compendium, which covers 376 cited publications from January through December. Topics range from bicycle/pedestrian studies to automated vehicle feasibility to energy consumption and greenhouse gas emissions.

2016 NHTS Pretest is Underway

On October 27, 2015, the Office of Management and Budget issued clearance to conduct the 2016 NHTS. This major milestone allows the 2016 NHTS to proceed with pretesting. The NHTS team anticipates that the first pretest respondent will be able to log into the online questionnaire by the first week of December. Pretesting began in mid-November with recruitment survey mail outs and will continue until January.

Are you a participant in the NHTS survey? If so, please visit the NHTS survey site. It has everything you need to get started.

Expert Panel Review Summary Report

A Panel of Expert Survey Methodologists met at the USDOT Headquarters on April 28, 2015 to review the 2016 NHTS Data Collection Plan, provide suggestions for improvement and discuss the future of the NHTS Program. The NHTS was fortunate to tap these world-renowned experts for measures of caution in the survey redesign and response analysis process. The summary report can be found here [PDF].

Stay Informed

Join the NHTS News mail list.

NHTS News is a light-traffic mail list with announcements and other information of interest to the NHTS Community.

Recent Publications

2015 NHTS Compendium of Uses, now through December.

CDC report Active Transportation Surveillance — United States, 1999–2012.

TRB NHTS Task Force mid-year meeting minutes [PDF].

TRB NHTS Task Force Phase One Report: Exploring New Directions for the National Household Travel Survey.

Security Notice | Plug-Ins

Developed by the CTA, ORNL under funding from the FHWA

National Household Travel Survey

The National Household Travel Survey (NHTS) provides information to assist transportation planners and policy makers who need comprehensive data on travel and transportation patterns in the United States.

What they do

The NHTS/NPTS serves as the nation's inventory of daily travel. Data is collected by surveying recent trips taken by household members, and includes:

- purpose of the trip (work, shopping, etc.);
- means of transportation used (car, bus, subway, walk, etc.);
- how long the trip took, i.e., travel time;
- time of day when the trip took place;
- day of week when the trip took place; and
- if a private vehicle trip:
- number of people in the vehicle , i.e., vehicle occupancy;
- driver characteristics (age, sex, worker status, education level, etc.); and
- vehicle attributes (make, model, model year, amount of miles driven in a year).

National Household Travel Survey

NHTS data are used to:

- quantify travel behavior,
- analyze changes in travel characteristics over time,
- relate travel behavior to the demographics of the traveler, and
- study the relationship of demographics and travel over time.

The NHTS data are used primarily for gaining a better understanding of travel behavior. The data enable DOT officials to assess program initiatives, review programs and policies, study current mobility issues, and plan for the future.

The transportation research community, including academics, consultants and government, use the NHTS extensively to examine:

- travel behavior at the individual and household level;
- the characteristics of travel, such as trip chaining, use of the various modes, amount and purpose of travel by time of day and day of week, vehicle occupancy, and a host of other attributes;
- the relationship between demographics and travel; and
- the public's perceptions of the transportation system.

Crowd-sensing

Crowd-sensing is a jargon term for crowd-sourced data collection, i.e. data acquisition done by combining readings from numerous sensors/devices carried by different individuals. May or may not require implicit interaction with users.

This approach:

- is cost effective
- does not require hardware installation and maintenance
- has variable and somewhat uncontrolled coverage, both spatially and temporally

Examples:

Google maps provide travel time estimates based on crowd-sourced data from android phones. What other applications can we build?

Can we measure speed? Detect accidents? Detect traffic volumes?
Can we build traffic light control system based on crowd-sourced data?

Issues with crowd-sensing

Sampling bias: data is collected in such a way that some members of the intended population are less likely to be included than others. Can take various forms:

non-response, preferential selection, spatial effects, network effects, …
(think of an example of each)

Why does it matter?
The data you collect may not be accurate in representing the whole population

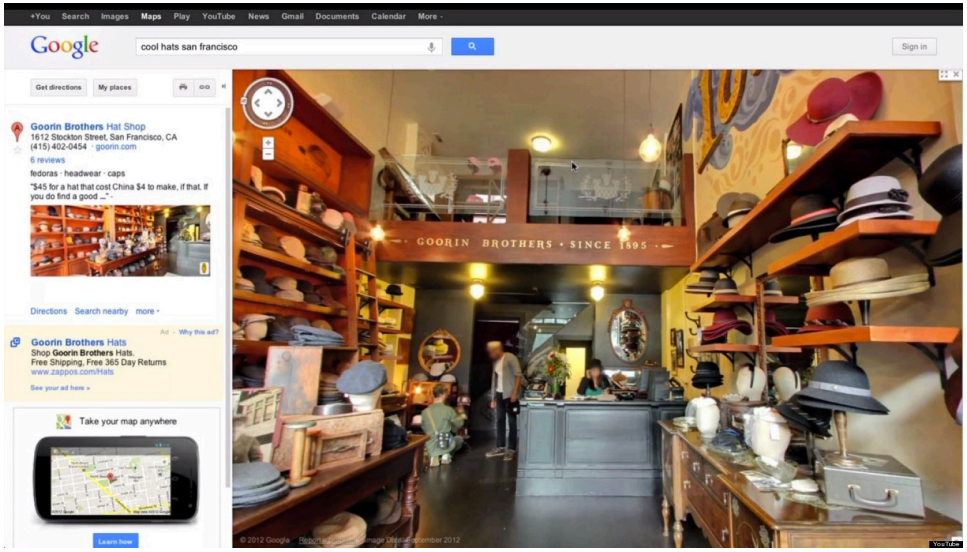
How can we know if it is the case?
By being thoughtful in comparing the characteristics of respondents to what you know about the population in general

What can we do about it?
Be careful about how we report the results. Be aware that the conclusions may only hold for a part of the population. Design crowd-sensing thoughtfully.

Who uses crowd-sourced data and crowd-sensing

WIKIPEDIA
The Free Encyclopedia

OpenStreetMap



Waze

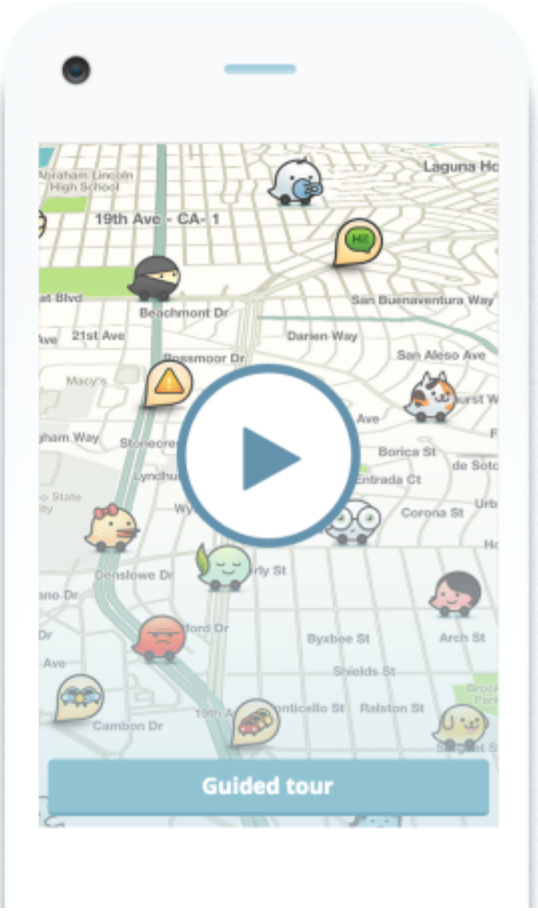
waze LIVE MAP MAJOR EVENTS SUPPORT BLOG ABOUT Login

Get the best route, every day, with real–time help from other drivers.

Waze is the world's largest community-based traffic and navigation app. Join other drivers in your area who share real-time traffic and road info, saving everyone time and gas money on their daily commute.

WAZE. OUTSMARTING TRAFFIC, TOGETHER.

GET IT ON Google play | Download on the App Store | Download from Windows Phone Store



SeeClickFix

SeeClickFix

PRODUCTS CUSTOMERS LEARN MORE REQUEST A DEMO

Products

- Mobile Apps for Citizens
- Mobile Apps for Officials
- Request Management
- Website Widgets
- Knowledge Base
- City Notifications
- Integrations and APIs
- Features and Pricing

Join our mailing list

SeeClickFix updates in your inbox!

Email*

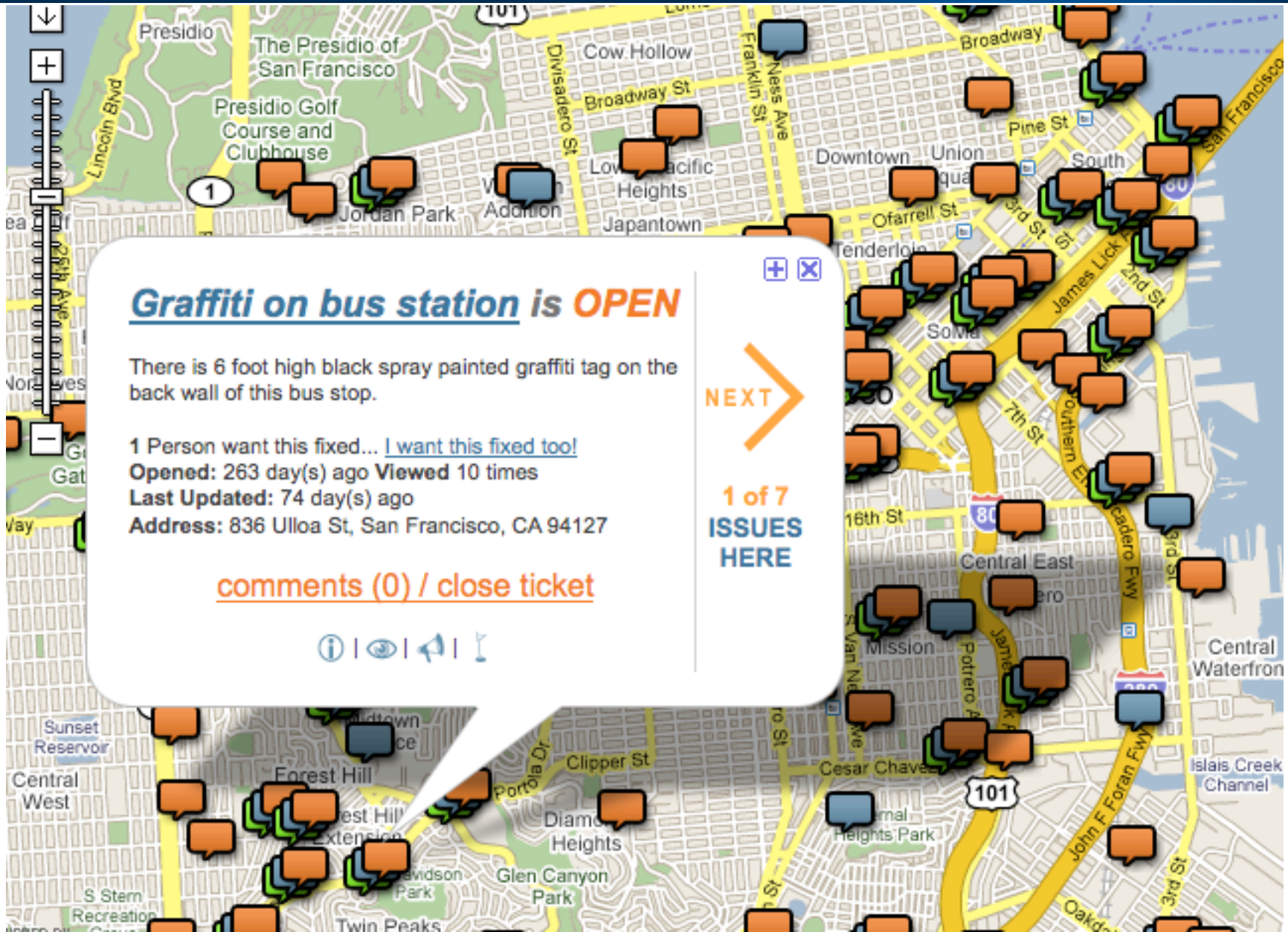
Subscribe

Mobile Apps for Citizens

City Branded Apps Connect Citizens

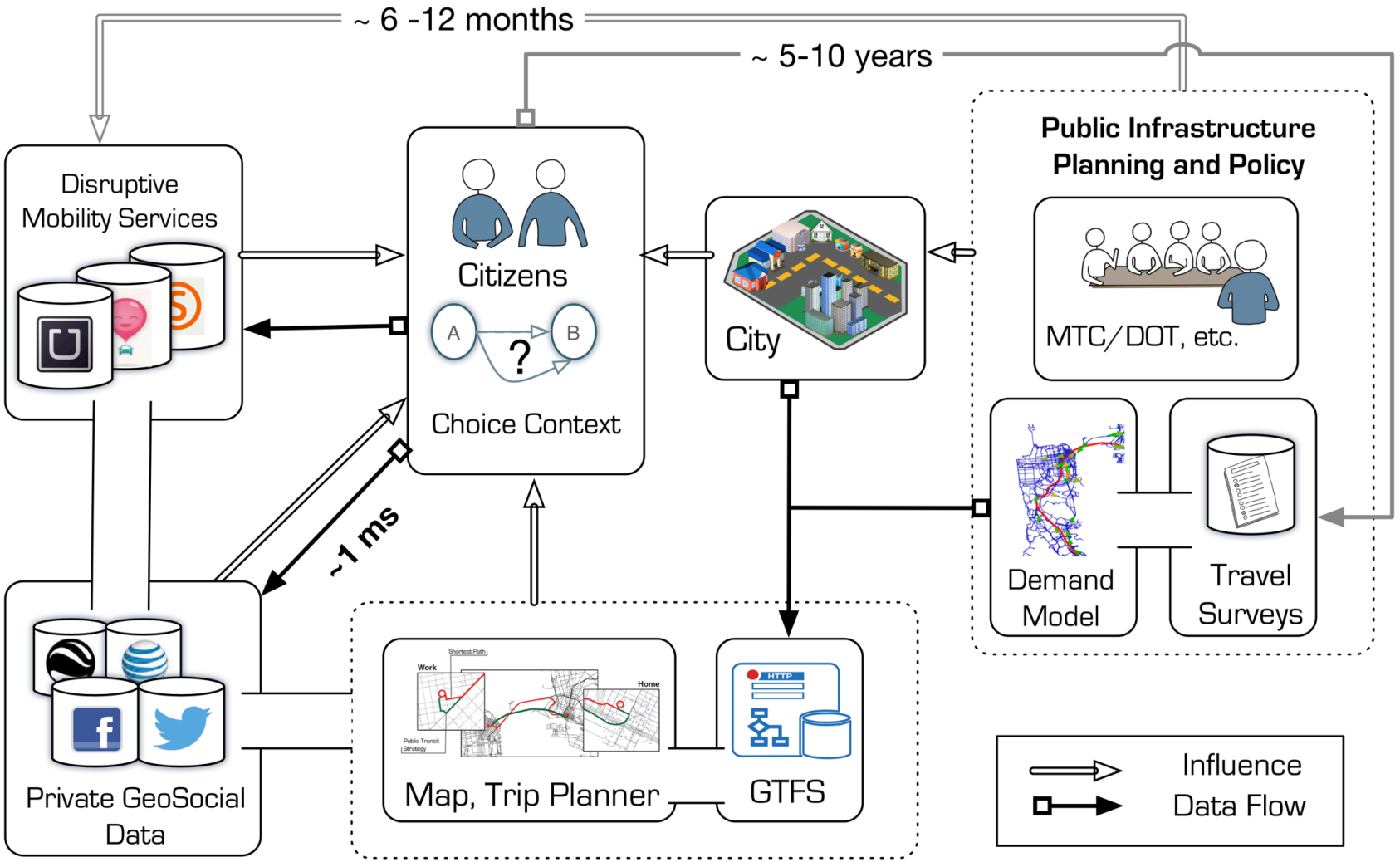
Innovation in the Palm of Your Hand

SeeClickFix



Data flows, ownership, exchange

Urban mobility data ecosystem



Open Data

https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

In general, open data will be consistent with the following **principles**:

Public. Consistent with Open Government Directive, agencies must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security.

Accessible. Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable.

Described. Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them.

Reusable. Open data are made available under an open license that places no restrictions on their use.

Complete. Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements. Derived or aggregate open data should also be published but must reference the primary data.

Timely. Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs.

Managed Post-Release. A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements.

Oakland

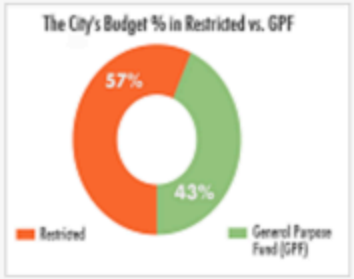
https://data.oaklandnet.com/

CITY OF OAKLAND

Sign Up Log In Help

Search

Home Data Getting Started Developers



Proposed Budget for Fiscal Years 2015-2017

On May first the Mayor

Financial Forecast 2016-2020

Every two years the city

CrimeWatch Maps Past 90-Days

We are committed to making

Public Works Service Requests

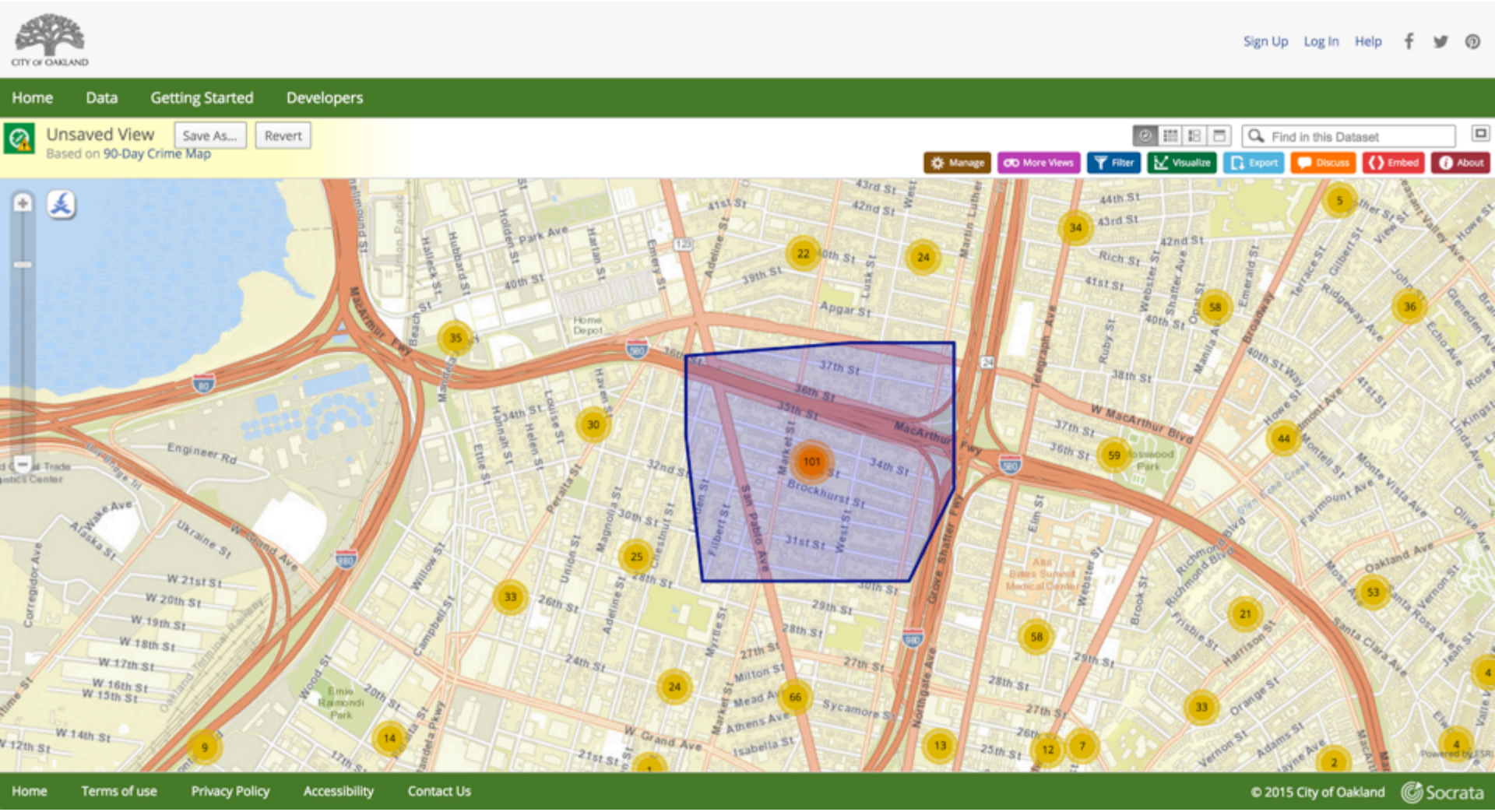
You can view public works

Resilience

Resilience is the capacity of individuals, communities and

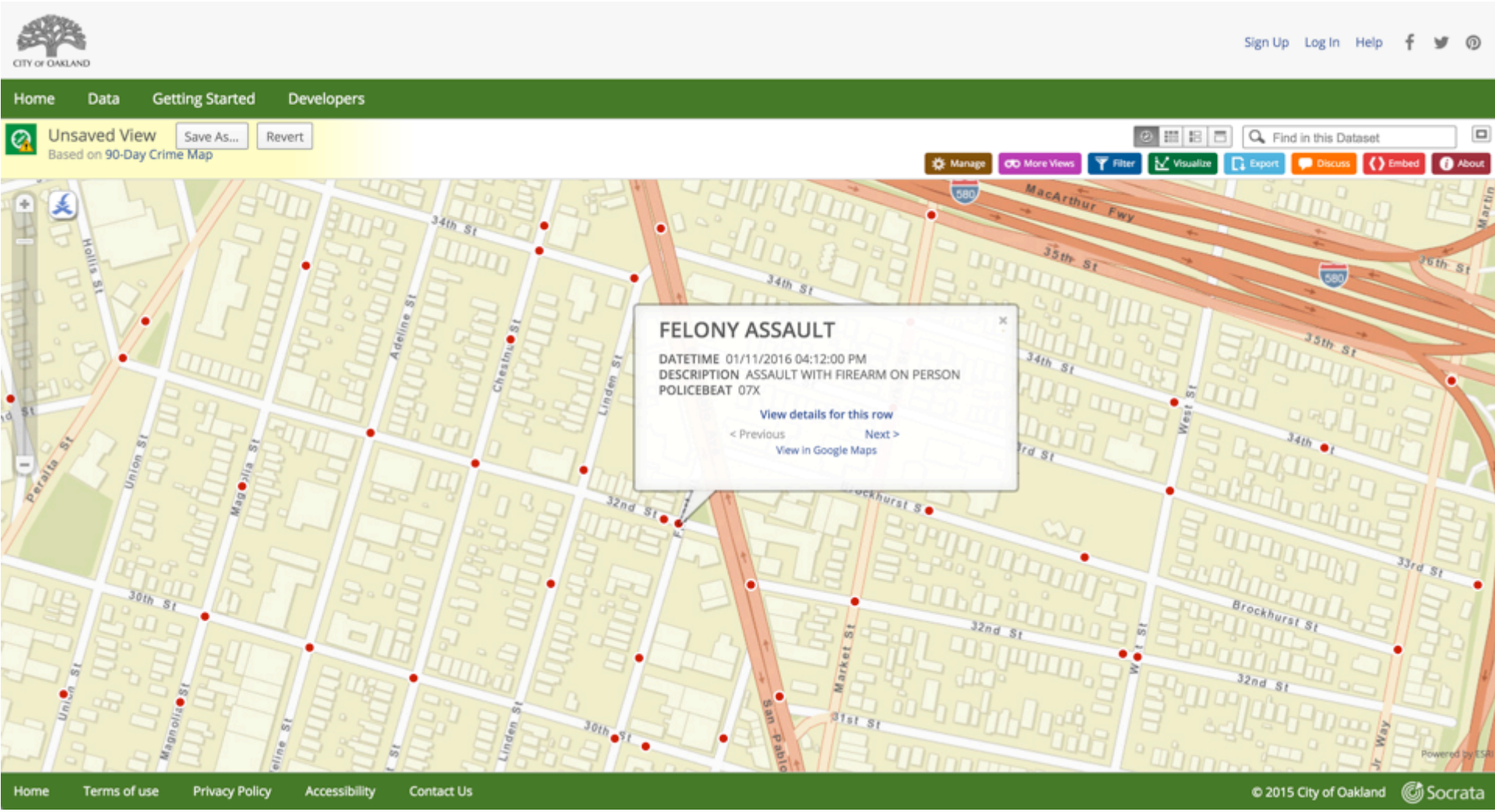
Oakland

90-Day Crime Map



Oakland

90-Day Crime Map



Berkeley

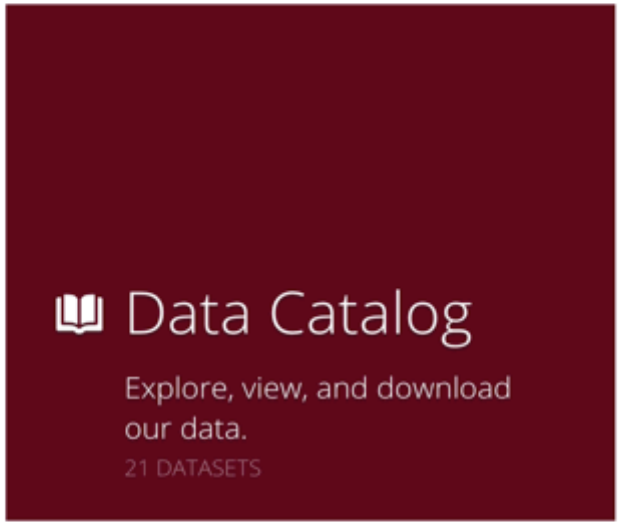
https://data.cityofberkeley.info/

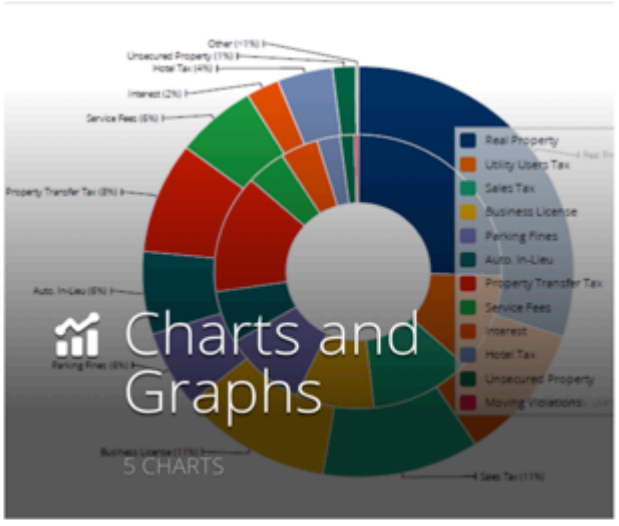
City of Berkeley Open Data

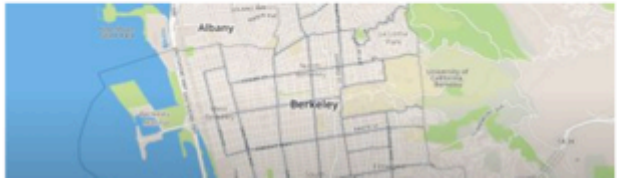
Catalog Developers Sign In

Search Data

Welcome to the City of Berkeley Open Data portal. Please use the brief (1 to 5 minute) tutorial videos below to help understand how to use the Open Data portal features. Please use the City's Online Service Request form to offer feedback or ideas regarding this site.

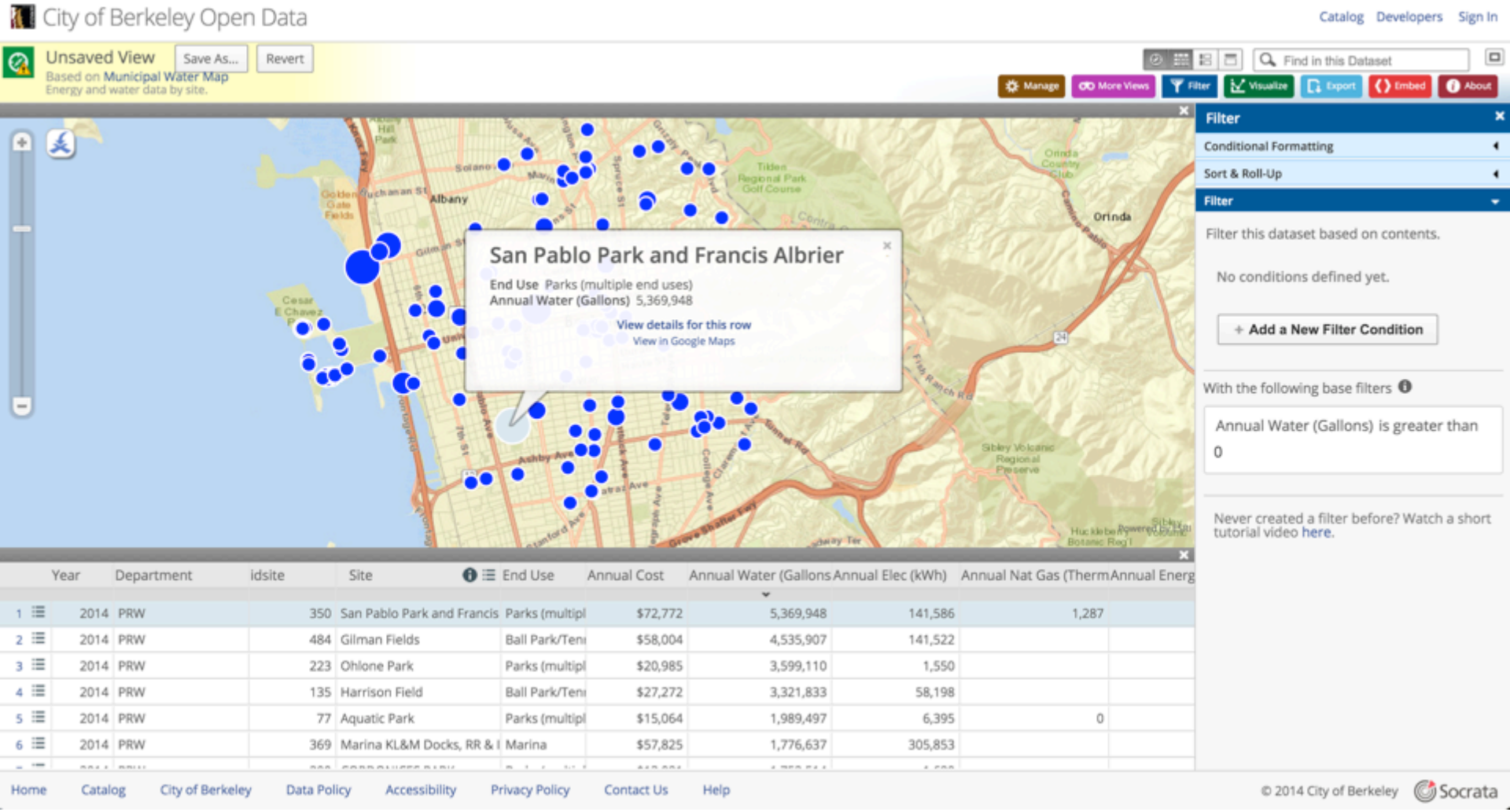






Berkeley

Municipal Water Map

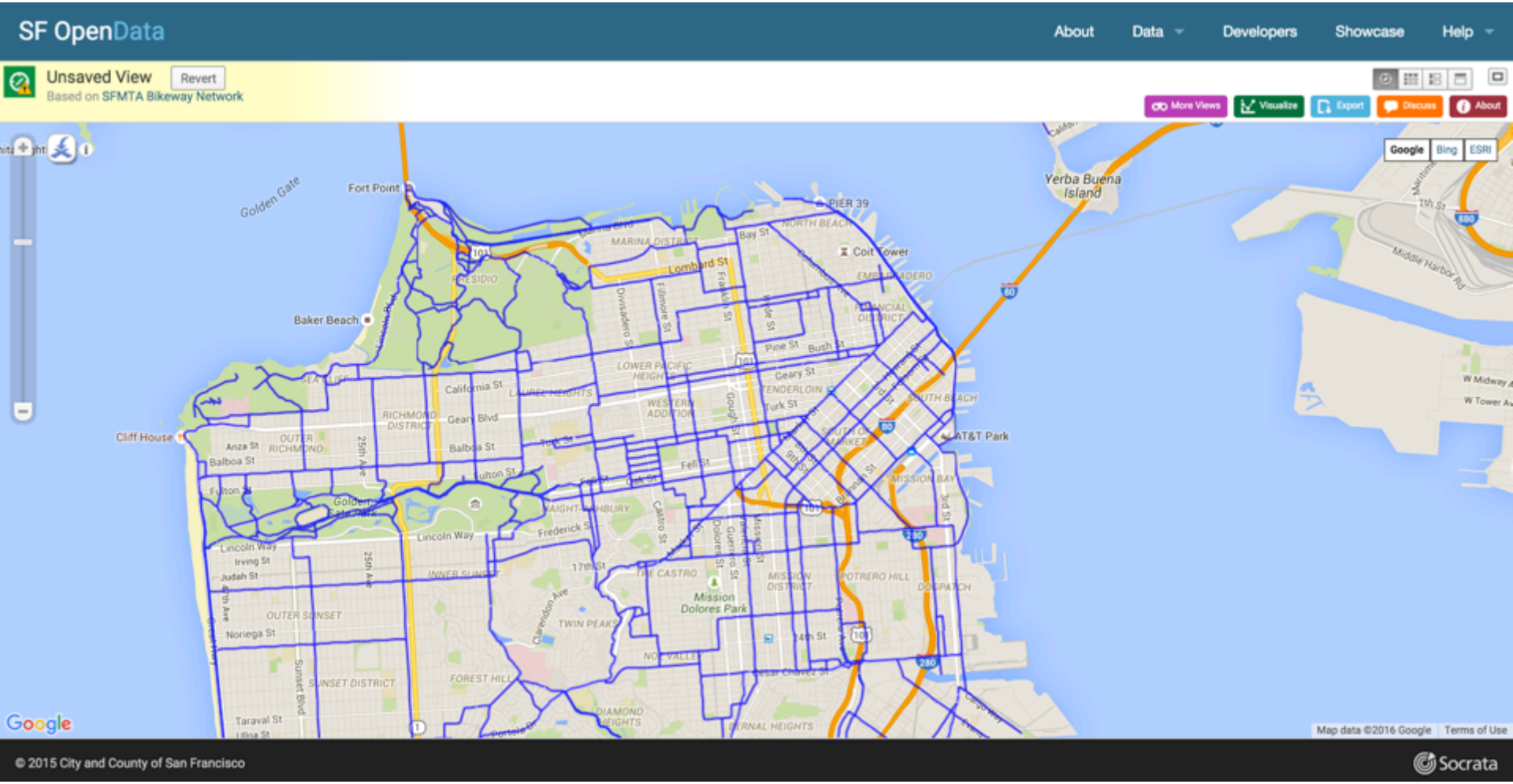


San Francisco

https://data.sfgov.org/

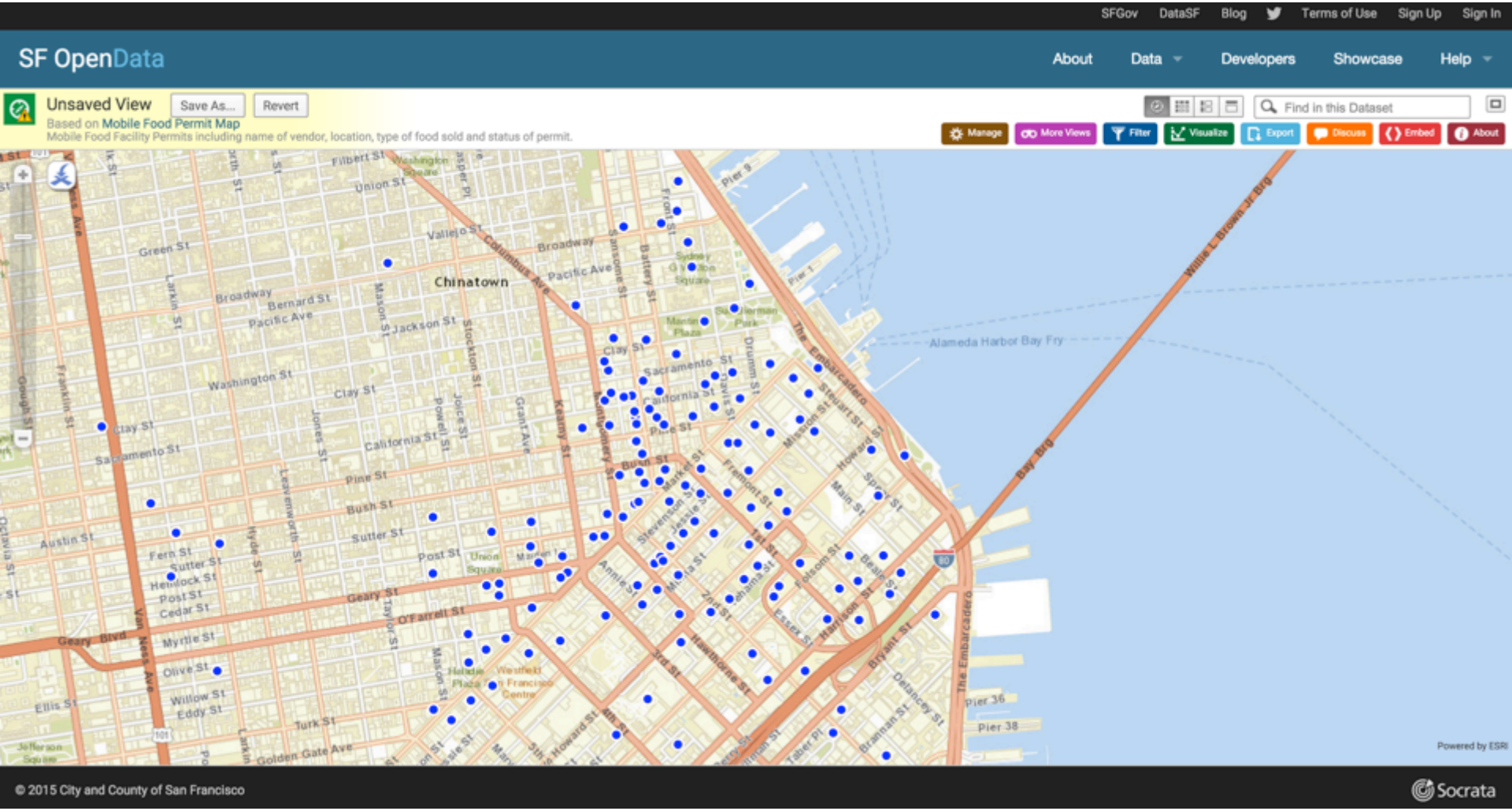
San Francisco

The network of bike routes, lanes, and paths around the city of San Francisco.



San Francisco

Mobile Food Permit Map

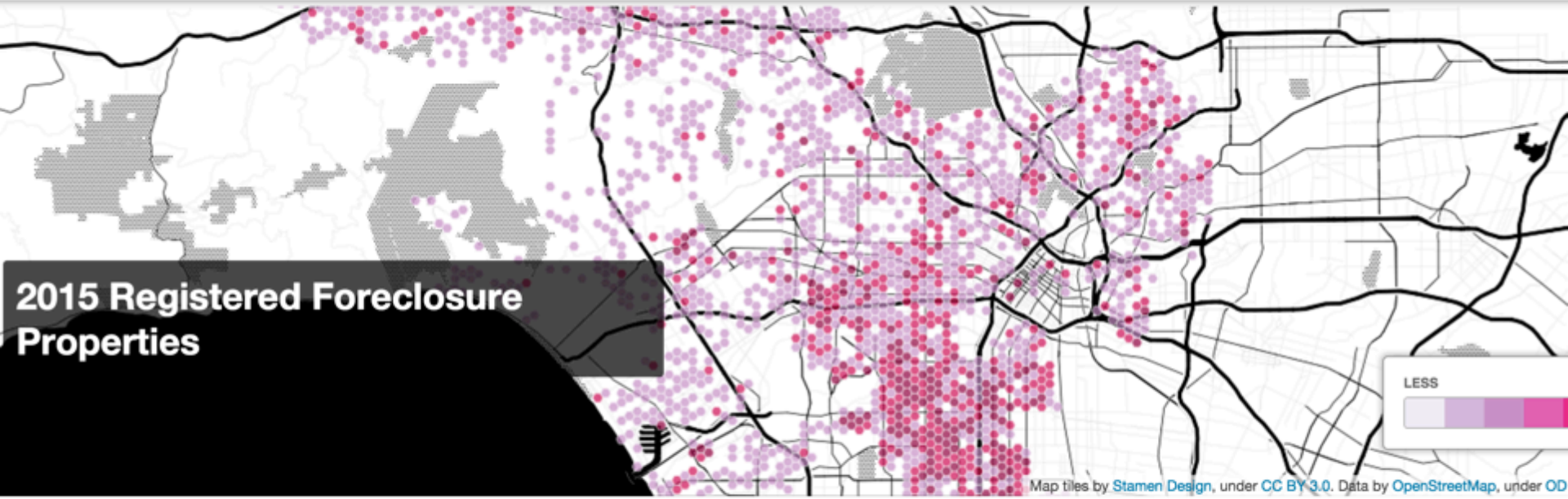


Los Angeles

https://data.lacity.org/

DATA L.A.

Data Catalog Developer Resources News & Updates Sign Up Log In



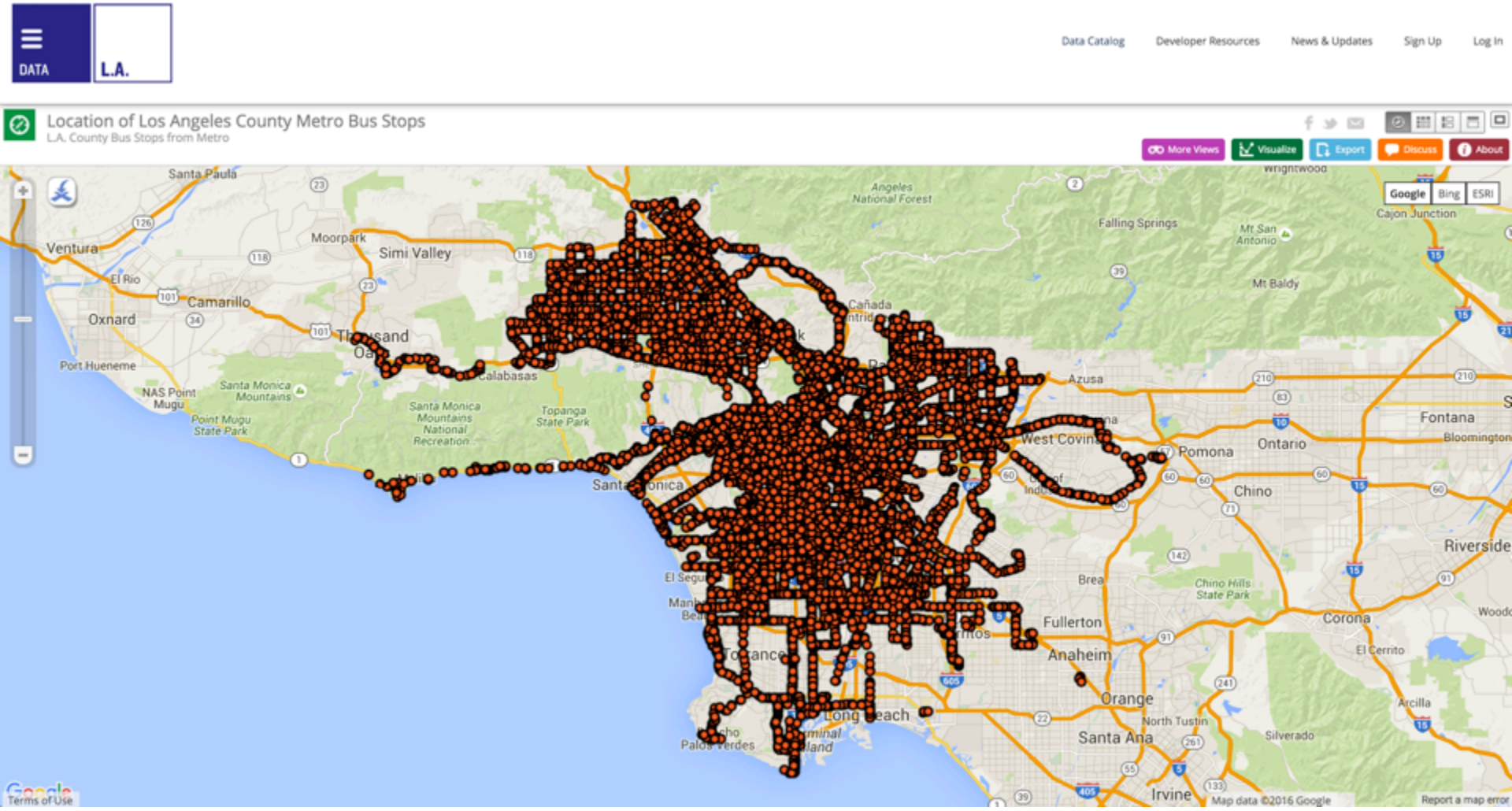
LA

LOS ANGELES OPEN DATA

Information, Insights, and Analysis from the City of Los Angeles

Los Angeles

Location of Los Angeles County Metro Bus Stops



Energy.Gov

http://energy.gov/data/articles/who-uses-open-data

ENERGY.GOV

Search Energy.gov

PUBLIC SERVICES | SCIENCE & INNOVATION | ENERGY SAVER | ABOUT ENERGY.GOV | OFFICES

Home » Who Uses Open Data?

Who Uses Open Data?

July 15, 2013 - 10:35am

Tweet G+1 Pin it

RELATED ARTICLES

The Booming App Economy

U-Launch Winner Secures $2.4M Investment for Building Energy Performance Software

President Obama Lauds Virginia Energy

APIs

An **application programming interface** (API) is a set of subroutine definitions, protocols, and tools for building software and applications.

An API specification may include specifications for routines, data structures, object classes, variables, or remote calls.

Web APIs are the defined interfaces through which interactions happen between an enterprise and applications that use its assets. When used in the context of web development, an API is typically defined as a set of Hypertext Transfer Protocol (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language (XML) or JavaScript Object Notation (JSON) format. While "web API" historically has been virtually synonymous for web service, the recent trend (so-called Web 2.0) has been moving away from Simple Object Access Protocol (SOAP) based web services and service-oriented architecture (SOA) towards more direct representational state transfer (REST) style web resources and resource-oriented architecture (ROA).[19] Part of this trend is related to the Semantic Web movement toward Resource Description Framework (RDF), a concept to promote web-based ontology engineering technologies. Web APIs allow the combination of

APIs

APIs are one of the most common ways technology companies integrate with each other. The main policies for releasing an API are:

Private: The API is for internal company use only.

Partner: Only specific business partners can use the API.

Public: The API is available for use by the public. An important factor when an API becomes public is its interface stability. Changes by a developer to a part of it—for example adding new parameters to a function call—could break compatibility with clients that depend on that API.

Data are often made accessible via a public API that describes request parameters and response formats.