

CITY OF #BIGDATA

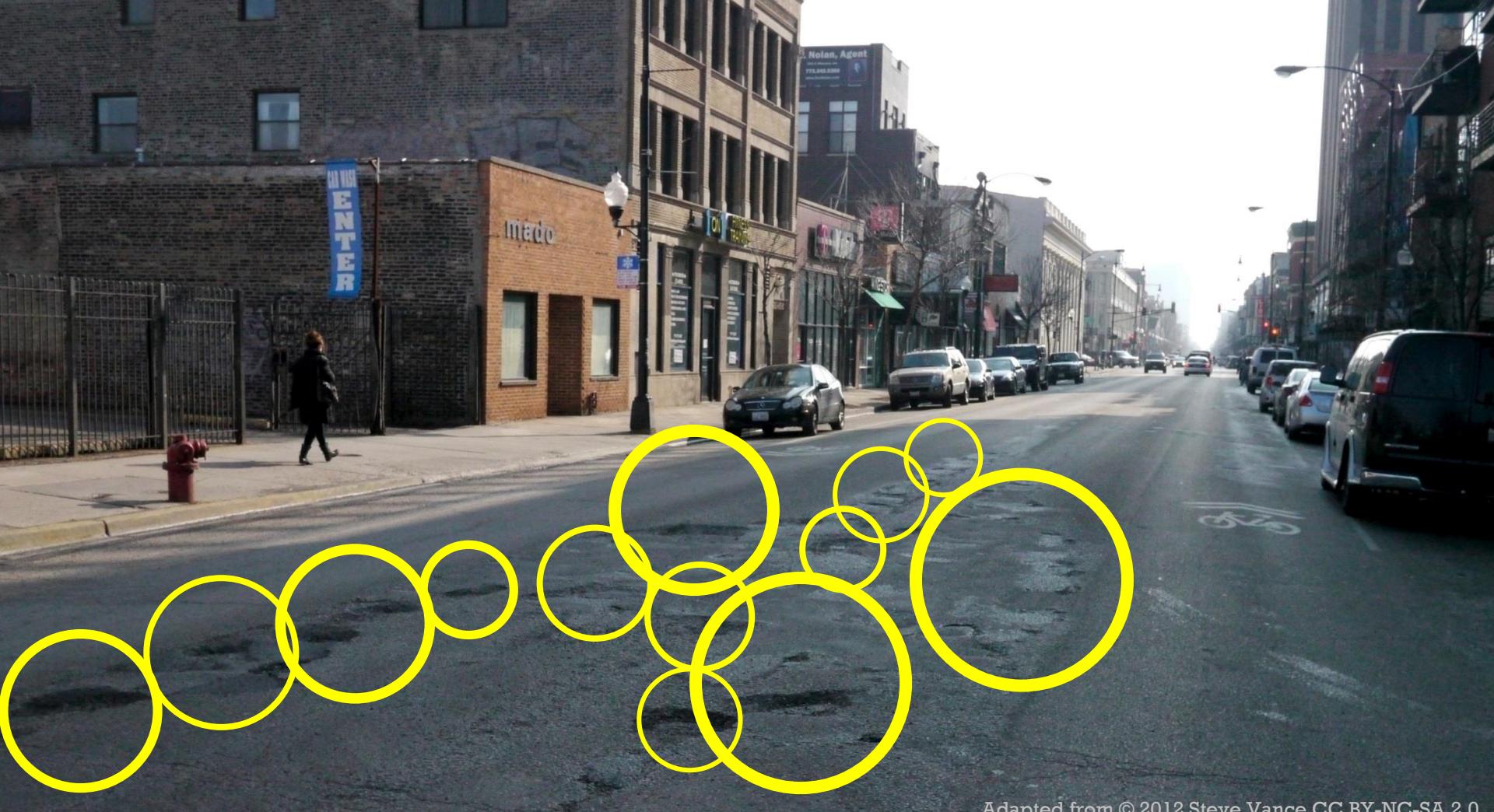
TOM SCHENK JR.
CHIEF DATA OFFICER, CITY OF CHICAGO
@CHICAGOCDO



**IN CHICAGO, WE BELIEVE THAT THE
POWER OF TECHNOLOGY IS DRIVEN
BY THE PEOPLE WHO USE AND
BENEFIT FROM IT.**



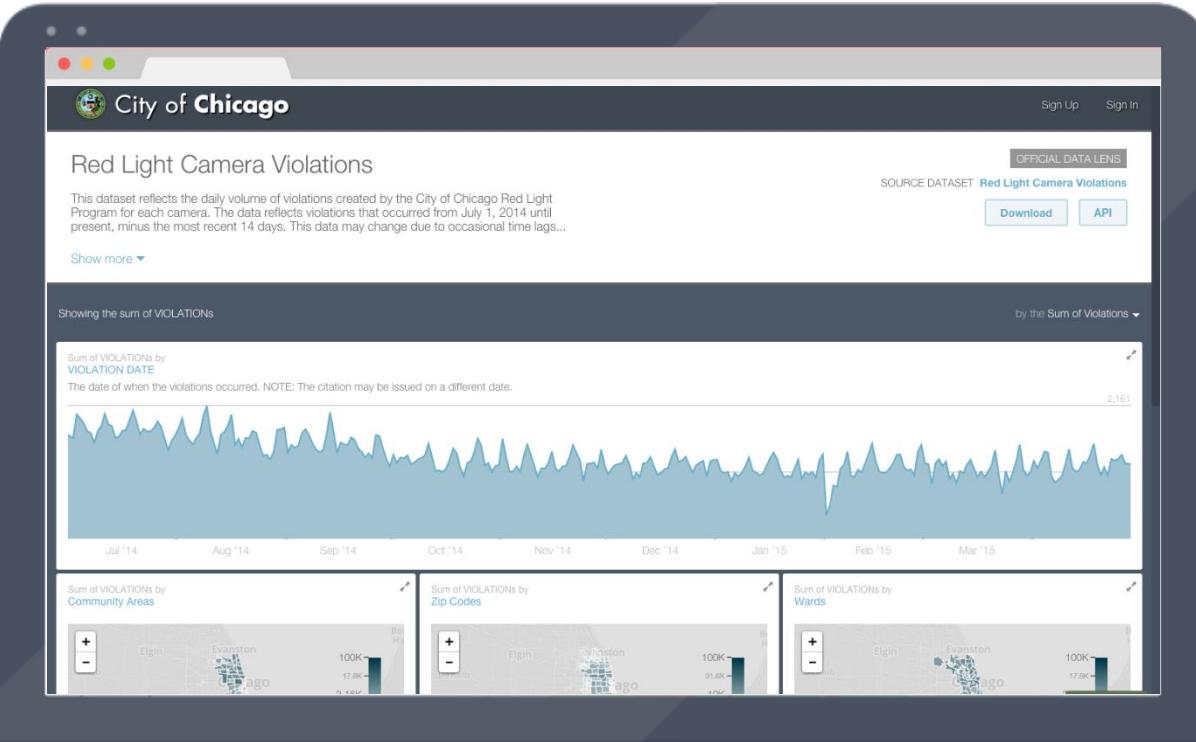
© 2012 _chrisUK CC-BY-ND 2.0



Adapted from © 2012 Steve Vance CC BY-NC-SA 2.0

	CREATION DATE	STATUS	COMPLETION DATE	SERVICE REQUEST NUMBER	TYPE OF SERVICE REQUEST	CURRENT ACTIVITY	MOST RECENT ACTION
1	11/14/2014	Open		14-01987391	Pothole in Street		
2	11/13/2014	Open		14-01985287	Pothole in Street		
3	11/13/2014	Open		14-01984936	Pothole in Street		
4	11/13/2014	Open - Dup		14-01985890	Pothole in Street		
5	11/13/2014	Open		14-01984723	Pothole in Street		
6	11/13/2014	Completed	11/13/2014	14-01985227	Pothole in Street	Final Outcome	Pothole Patched
7	11/13/2014	Open		14-01964083	Pothole in Street		
8	11/13/2014	Open		14-01963203	Pothole in Street		
9	11/13/2014	Open		14-01964892	Pothole in Street		
10	11/13/2014	Open		14-01963403	Pothole in Street		
11	11/13/2014	Open		14-01964109	Pothole in Street		
12	11/13/2014	Open		14-01963369	Pothole in Street		
13	11/13/2014	Open		14-01964355	Pothole in Street		
14	11/13/2014	Open - Dup		14-01963810	Pothole in Street		
15	11/13/2014	Open		14-01964366	Pothole in Street		
16	11/13/2014	Open		14-01964183	Pothole in Street		
17	11/13/2014	Open		14-01964745	Pothole in Street		
18	11/13/2014	Completed	11/13/2014	14-01965383	Pothole in Street	Final Outcome	Pothole Patched
19	11/13/2014	Open		14-01964782	Pothole in Street		
20	11/13/2014	Open		14-01960594	Pothole in Street		
21	11/13/2014	Open		14-01984378	Pothole in Street		
22	11/13/2014	Open		14-01985179	Pothole in Street		
23	11/13/2014	Open		14-01984870	Pothole in Street		
24	11/13/2014	Open		14-01983495	Pothole in Street		
Total:		323428					

Data on potholes are reported by residents and city staff through the 311 system, which is then reported on the City's #opendata portal—updated daily.



Chicago has released more #opendata, including important items such as red light and speed camera violations, problem landlords, and public chauffeurs.

the City of Chicago data set

b. Nothing in third party that establishes adopting or implementing sets housed on the data portal to impair or damage the use of the data portal being used by the purpose; or (iv) terminate a data provided on the data set pursuant to Section 9(b) of applications based on City of the public data set, and

c. Nothing in enforce any provision of if result in any liability to the party that establishes or mu

12. Enforcement Standards

The chief information regulations to implement it

13. Effective Date

This Order shall take

RAHM EMANUEL
MAYOR



OFFICE OF THE MAYOR
CITY OF CHICAGO

RAHM EMANUEL
MAYOR

December 10, 2012

Susana A. Mendoza
City Clerk
Room 107, City Hall
Chicago, IL 60602

Dear Ms. Mendoza:

I transmit herewith for filing Executive Order No. 2012-2, which I have signed this date.

Your prompt attention to this matter is appreciated.

Sincerely,

Mayor

2012-01-05
SUSANA A. MENDOZA
CITY CLERK
CITY OF CHICAGO

Received and filed

L. RAHM EMANUEL, M

1. Definitions

"Chief data officer" means Technology or the chief data officer

"Chief information officer"

Executive Order 2012-2

In 2012, Chicago issued an executive order which formalized the open data portal, endowed powers to the Chief Data Officer, created an advisory committee to advise on the expansion of new datasets, and required an annual open data report.



LEVERAGE DATA AND NEW
TECHNOLOGY TO MAKE
GOVERNMENT MORE EFFICIENT,
EFFECTIVE, AND OPEN

INCREASE & IMPROVE CITY DATA

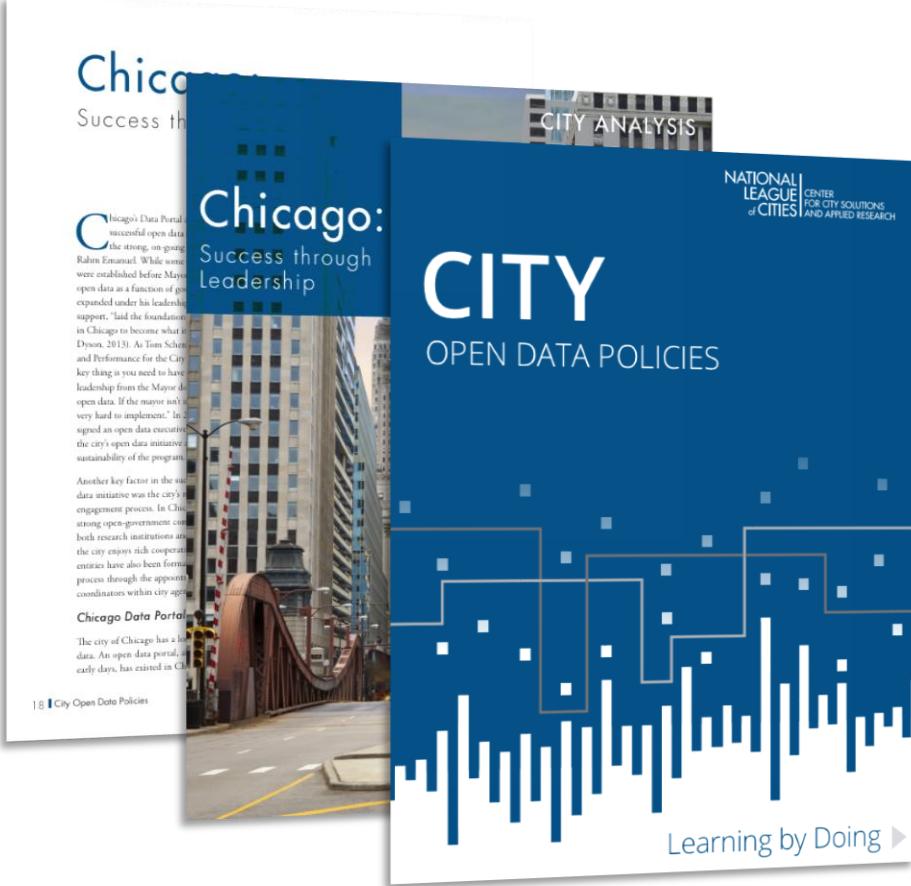
The City will continue to increase and improve the quality of City data available internally and externally, and facilitate methods for analyzing that data to help create a smarter and more efficient city.

**#OPENDATA PROVIDES A
MEANS TO CREATE AN
ECOSYSTEM AROUND DATA,
WHICH INCLUDES MULTIPLE
STAKEHOLDERS AND
INITIATIVES THAT EXTEND
BEYOND TRANSPARENCY.**

Open Gov Hack Night

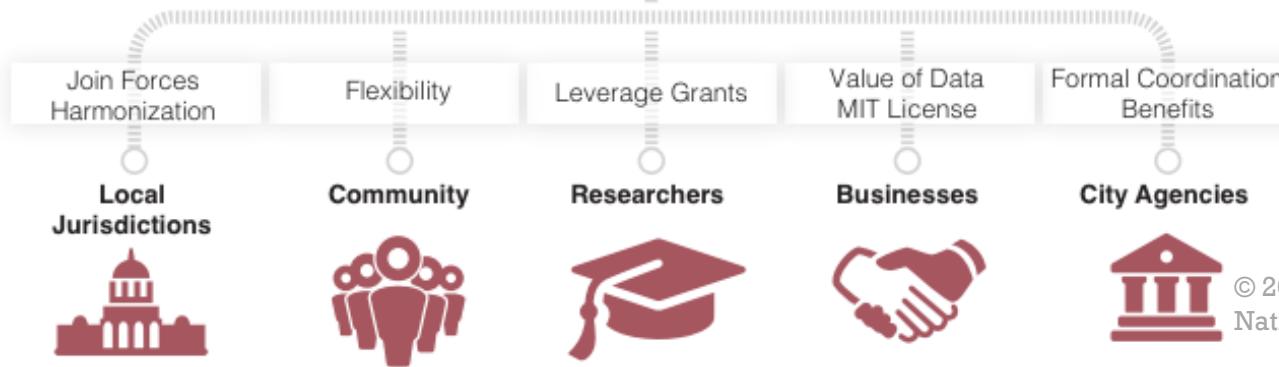
happy 1 year!





National League of Cities

NLC issued a report discussing the role of Chicago's leadership in developing a leading #opendata portal. The first chapter reviews Chicago's open data program and its benefits to the city, residents, and others.



© 2012
National League of Cities

“Open data initiatives are an increasingly popular component of governance. At the national level, Chicago’s #opendata initiative has been held up as a model for cities that are seeking to start their own open data programs.”

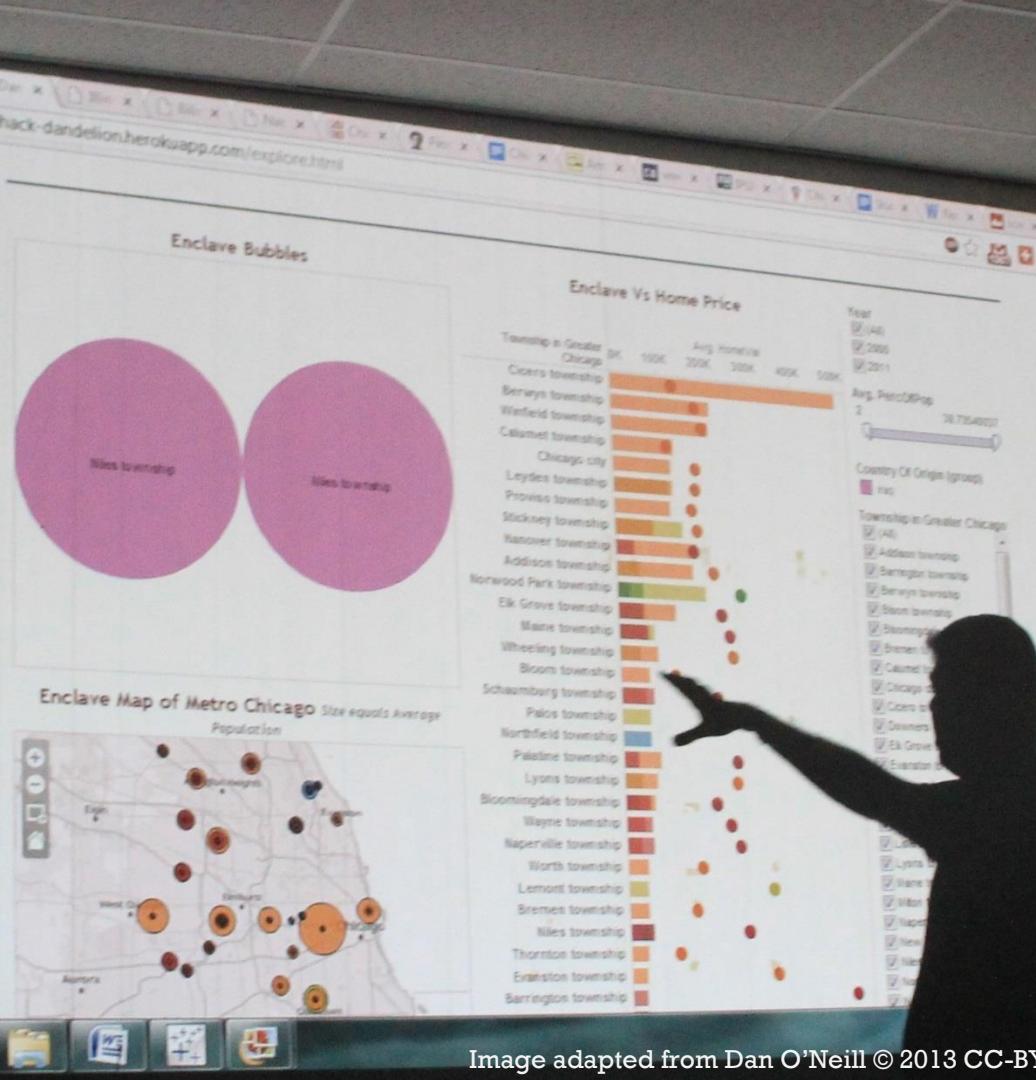
- National League of Cities, p. 22



© Tom Schenk Jr, 2016. CC-BY

Civic Tech Community

Chicago has a large, vibrant, **productive**, civic community. This is led by Chicago residents interested in technology and society. Smart Chicago Collaborative and non-profits provide assistance and city officials regularly engage in meetups and other activities. This group has produced several helpful apps using #opendata.



ChiHackNight

Meets every Tuesday
chihacknight.org

City Data User Group

Meets first Weds of month
meetup.com/ChicagoCityData

SWEETAROUND US

e.g. 3324 North California Avenue

Ward 42, Sweep Area 2

ARE THE DATES OF YOUR NEXT STREET SWEEPING

Want to be reminded of upcoming street sweepings?

SweepAroundUs has several options for being notified of street sweeping in this area: we can send you an email, a text message or you may subscribe to a calendar feed.¹

Enter your email address, and we'll send alerts directly to your inbox.

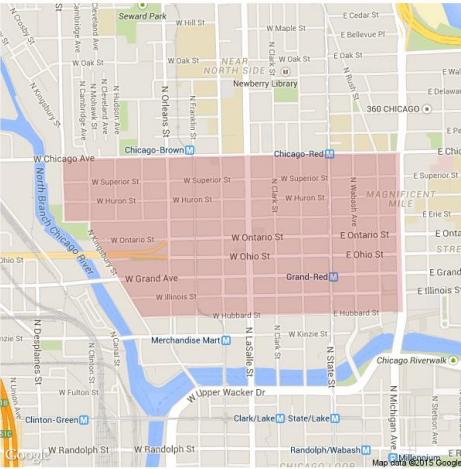
Receive text message alerts on your phone. To subscribe, text **W42A2** to **(773) 675-1985**.²

Download or subscribe to this area's sweep calendar in iCal, Google Calendar or Outlook.

2015 Sweep Schedule

1. Email and text message alerts will be sent one week, and one day, prior to the beginning of street sweeping in this area.

2. Standard message and data rates may apply. To unsubscribe from alerts for this area, text STOP W42A2.



sweetaround.us

Using #opendata, this service developed by the civic community alerts individuals to street sweeping activity by providing email, text, or calendar alerts.

\$1 LARGE LOTS

The City of Chicago is selling vacant residential lots for **\$1**. Here's how you get one.

Large Lots 3: Austin Ended on January 31st, 2015

The third Large Lot Program ended in Austin on **January 31st at 11:59 p.m.** Applicants will be contacted via a letter regarding the status of their application.

Check back later for updates!

Large Lots 2: East Garfield Park Ended on August 4th, 2014

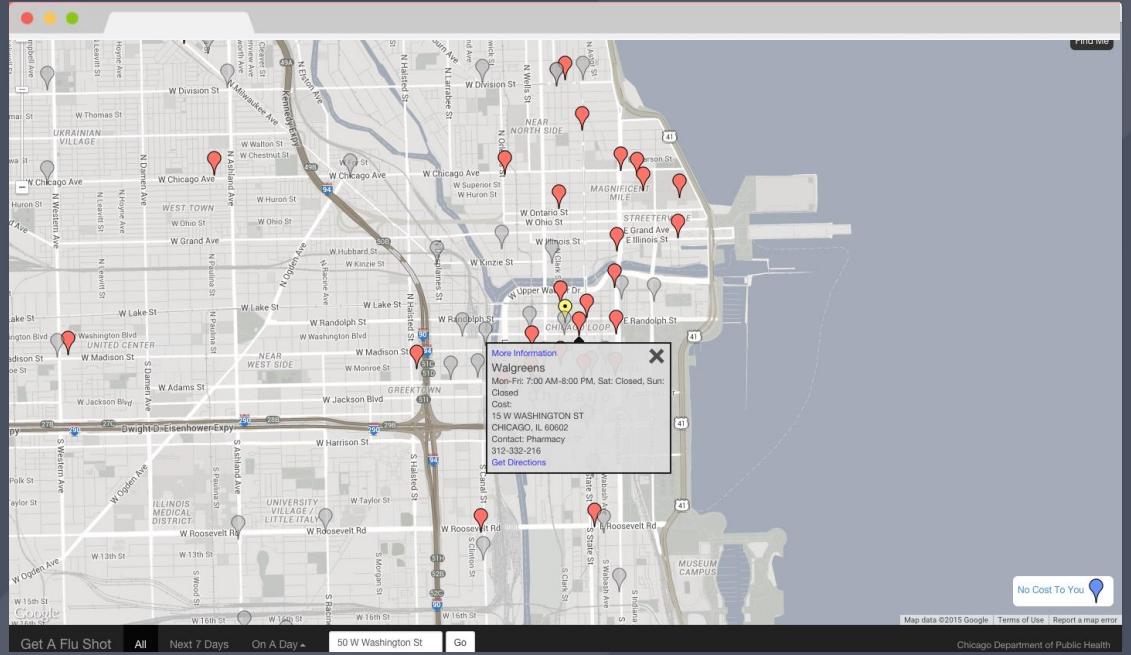
The second Large Lot Program ended in East Garfield Park on **August 4th at 11:59 p.m.** and 190 lots were successfully applied for. [Check the status of each lot](#)

[Check application status](#)

Check back later for updates!

largelots.org

The City of Chicago partnered with developers to create LargeLots, a website using #opendata to help residents apply to the City of Chicago \$1 lot program designed to encourage investment in struggling neighborhoods.

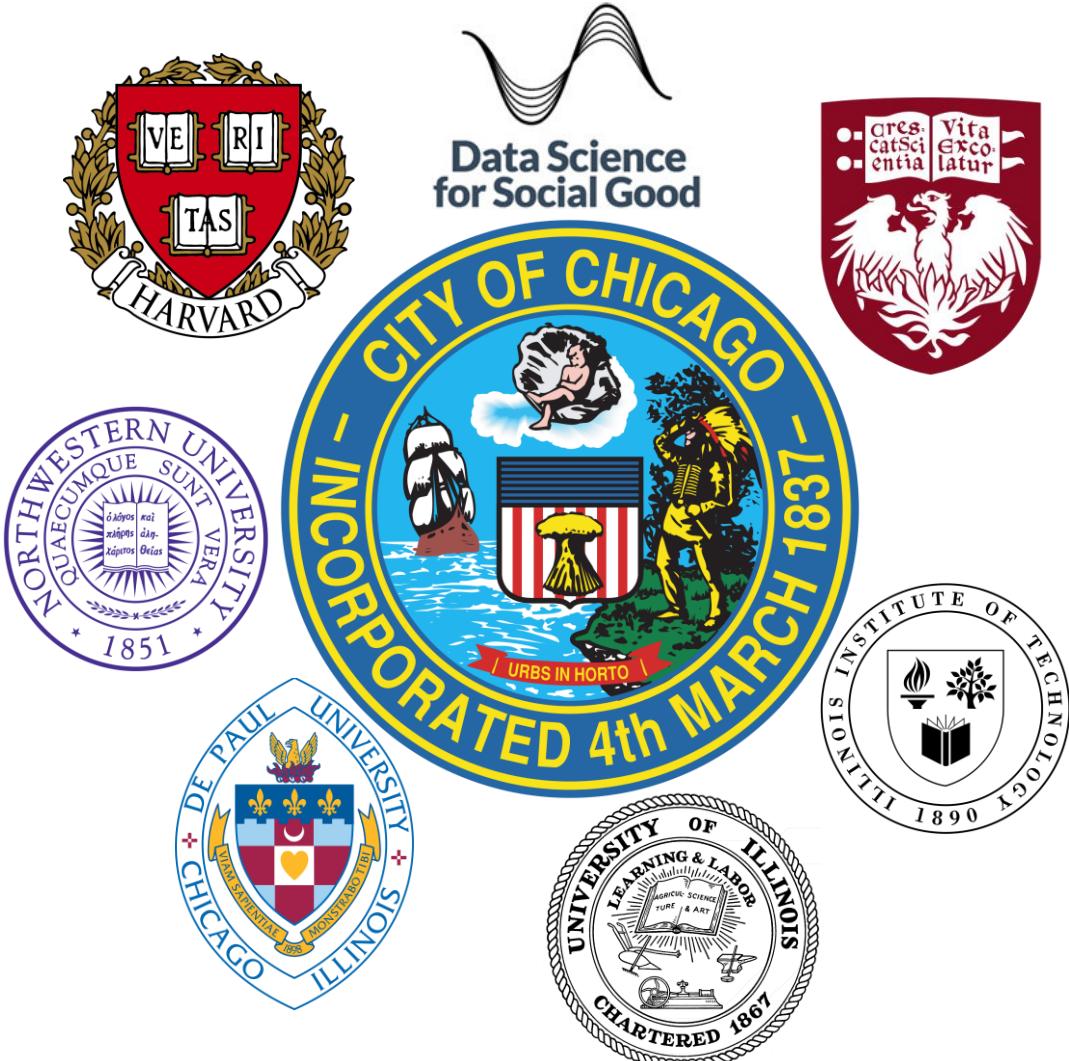


chicagoflushots.org

Chicago Flu Shots
was developed to
easily find flu-shot
locations across
Chicago. The code
was created by a
volunteer is
#opensource so the
site was adopted by
Boston, Philadelphia,
and San Francisco.

Academia

The City of Chicago has a number of high-quality research universities and groups willing to engage in projects with the city. We can leverage open data portal and data itself to create cooperative relationships.

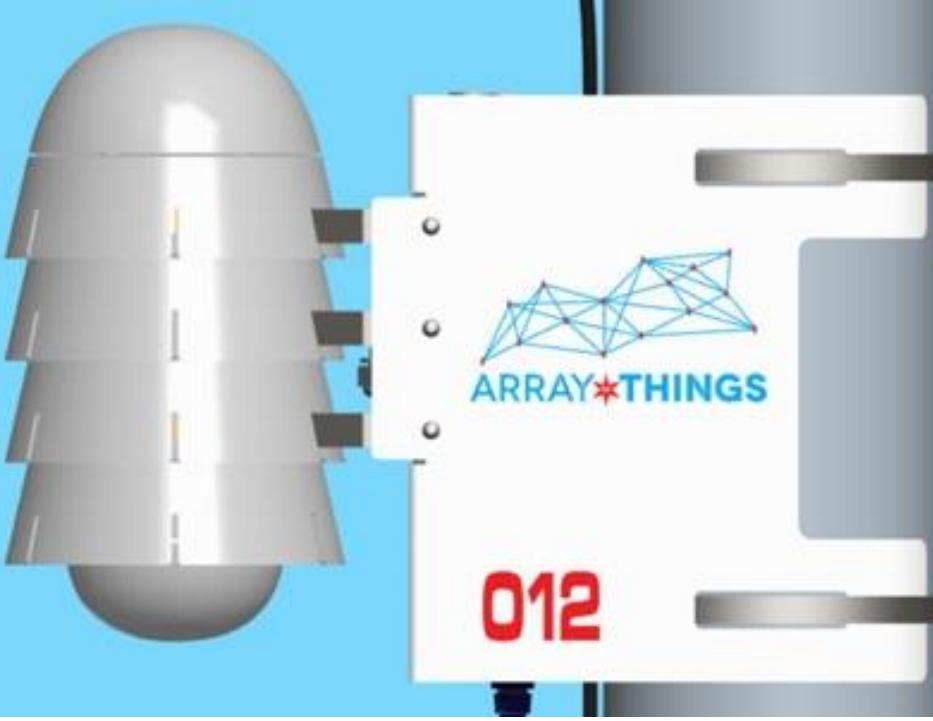


OPEN DATA & INTERNET OF THINGS



© 2015 Chris Metcalf

Open data has also spread to physical devices.
@chrismetcalf used traffic congestion data from the open data portal to generate an imp to provide a red or green light to denote heavy or light traffic congestion.



© 2015 University of Chicago

Array of Things

arrayofthings.github.io

University of Chicago has partnered with multiple institutions to build a mesh network of small sensors, dubbed the **Array of Things**, that will frequently post data for public consumption.

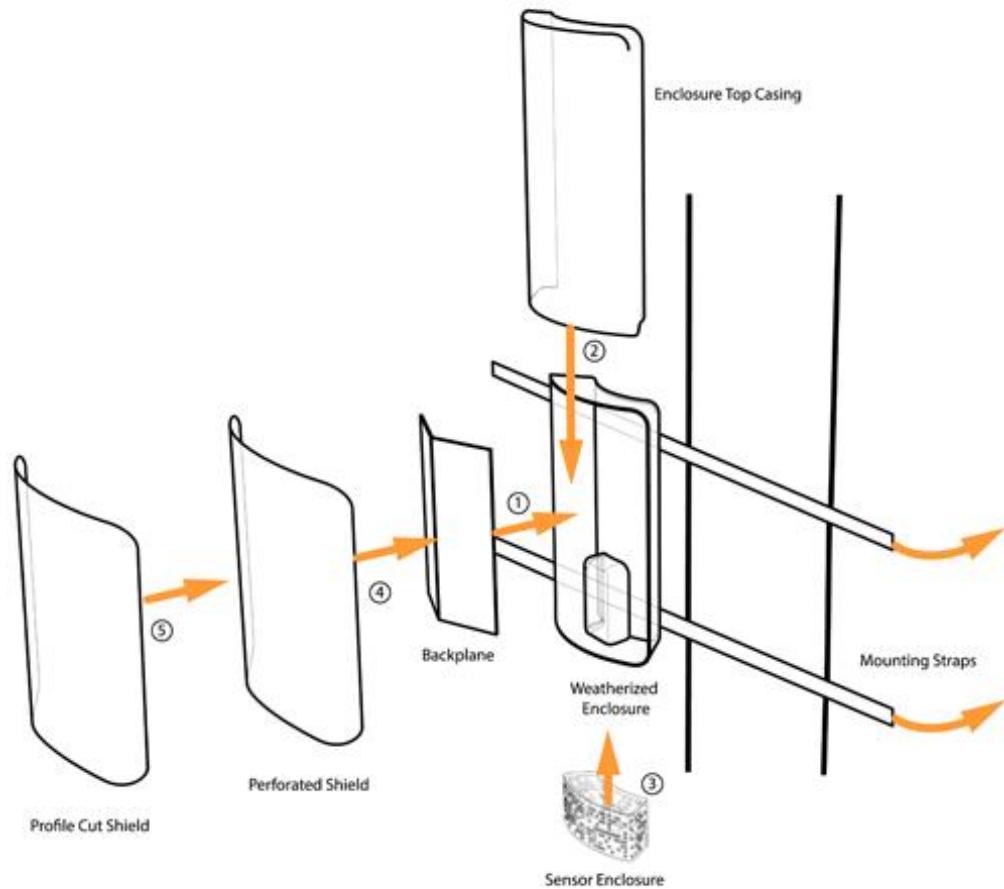
Array of Things



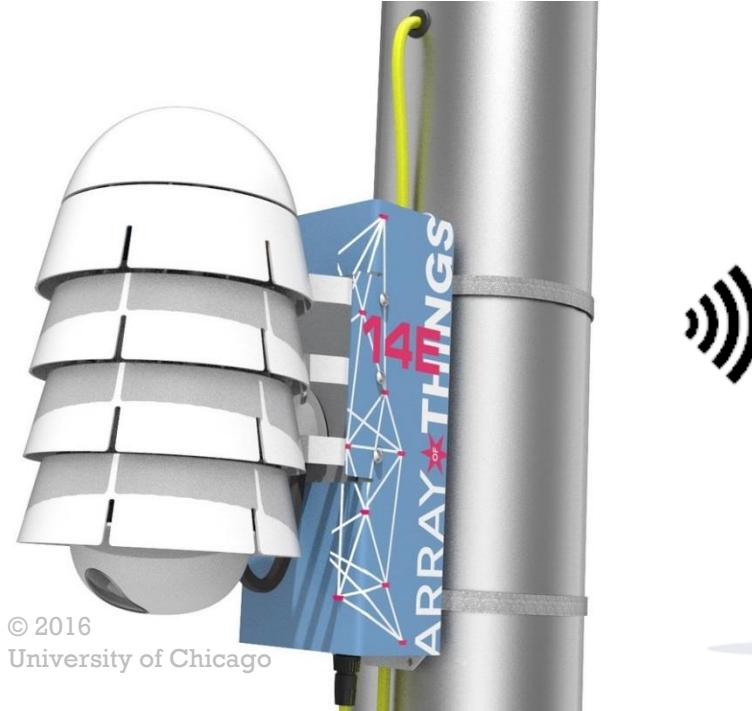
The Array of Things will provide hyper-local, temporal data on using a variety of sensors:

- Sensors measuring **sound and vibration**
- **Low-resolution infrared cameras** measuring sidewalk temperature
- **Climate and environmental data**, such as air-quality and temperature

Array of Things



OPEN INTERNET OF THINGS



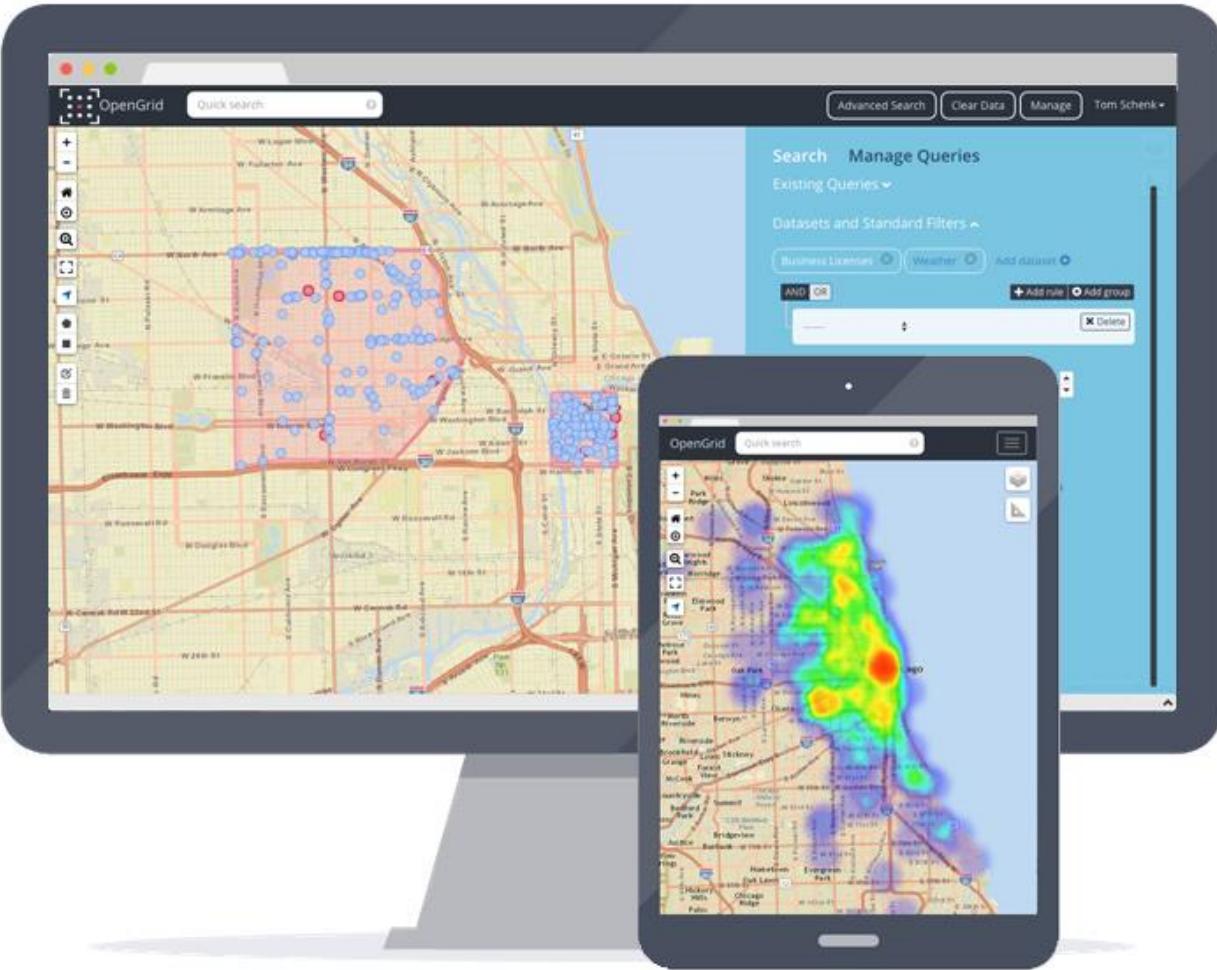
Array of Things

Name	Popularity	Type
Current Employee Names, Salaries, and Position Titles	111,229 views	Administration & Finance, personnel
Problem Landlord List - Map	147,000 views	Business, buildings, violence
Affordable Rental Housing Developers	148,631 views	Business, buildings, violence
Building Permits	202,232 views	Buildings, permits
Building Code Scofflaw List - Map	65,350 views	Business, buildings, violence
Business Licenses - Current Active	106,352 views	Business, licenses, permit
Crimes - 2001 to present	113,171 views	Public Safety, crime, police
Food Inspections - 2010 to present	118,598 views	Business, food safety, inspection
Chicago Street Names	94,406 views	Transportation, streets, geospatial
Police Stations - Locations, hours, facilities	135,642 views	Police, locations, facilities, geospatial
Libraries - Locations, Hours and Contact Information	93,172 views	Education, libraries, facilities, geospatial

Chicago Open Data Portal







OpenGrid.io

An #opensource platform which allows you to explore events such as 311 calls, crimes, permits, inspections, DIVVY trips in an interactive map. This software can be used by the public and an internal version drives situational awareness.

ADVANCED ANALYTICS

#Prediction

Using data and advanced research techniques to forecast and predict events in the city.

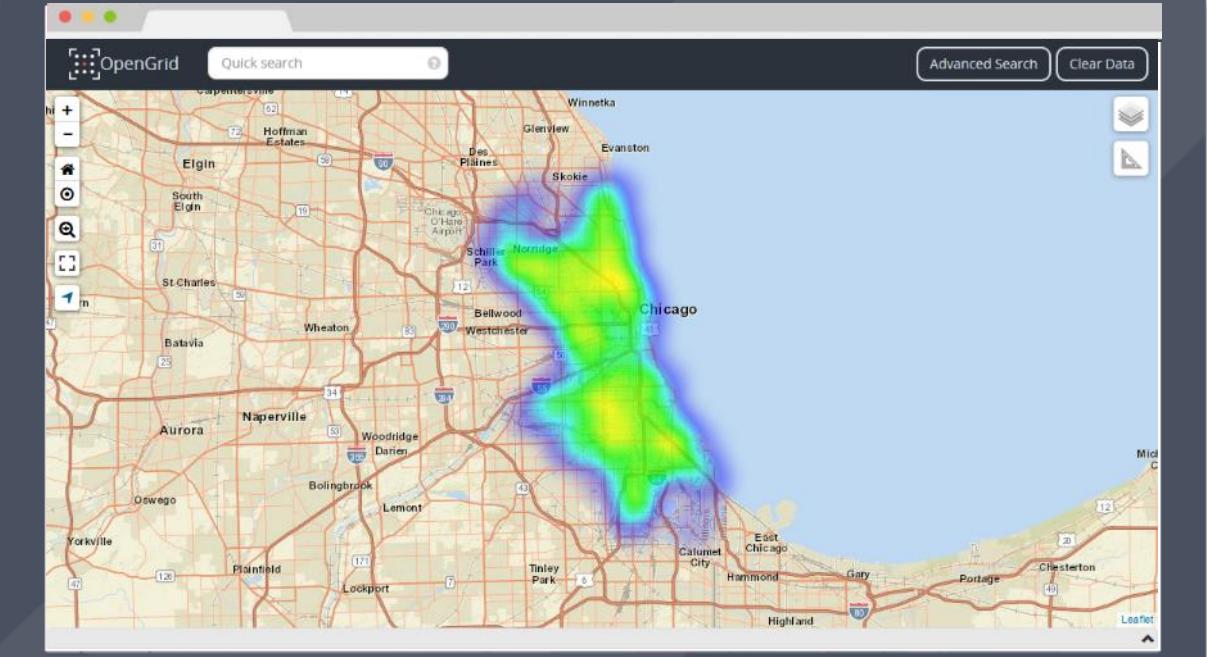
#Evaluation

Evaluate the effectiveness of programs, including the effectiveness of advanced analytics.

#Optimization

Optimizing the allocation of resources across the city to engage

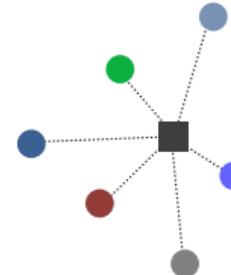




A map of rodent complaints across the city.

City of Chicago found 31 factors that predicted when and where rodent complaints are most likely in the next week. We used spatial-temporal relationships to create these #predictions, which started as an investigation of over 350 different factors.

Spatial Correlation



Temporal Correlation

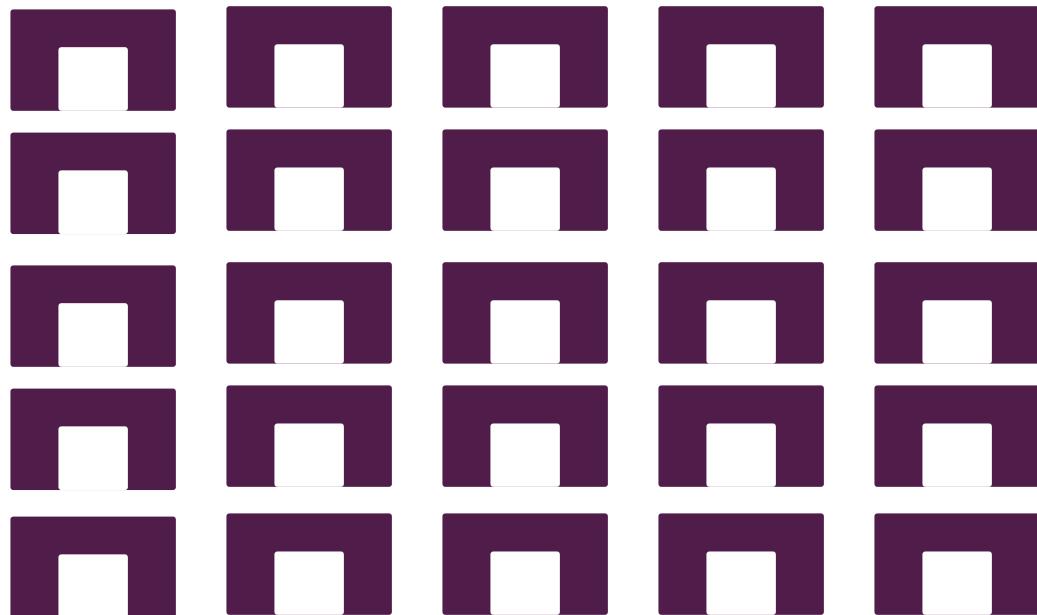
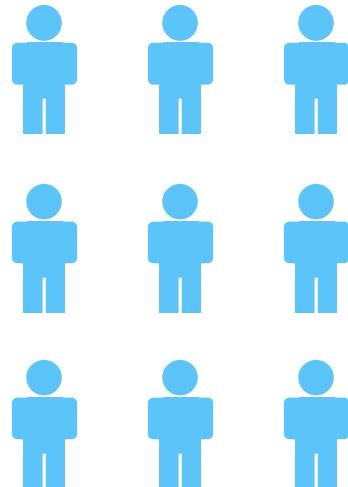


Address	Block	Score	X	Y	District
2514_W_BIRCHWOOD_AVE	202001000	0.00581066	1158200	1949710	24
3029_W_GRANVILLE_AVE	207005005	0.00517838	1154910	1940950	24
6982_N_CLARK_ST	107001013	0.00428247	1163450	1946340	24
6327_N_ROCKWELL_ST	208001003	0.00409735	1157880	1942020	24
6306_N_HERMITAGE_AVE	303001007	0.00393258	1163590	1941760	24
2905_W_GRANVILLE_AVE	207002001	0.0300252	1155830	1940980	24
6226_N_FAIRFIELD_AVE	208008005	0.0125533	1156830	1941340	24
2923_W_ROSEMONT_AVE	207006008	0.012348	1155830	1941820	24
6317_N MOZART_ST	207001001	0.0114882	1156230	1941870	24
6052_N_FAIRFIELD_AVE	208007005	0.0104587	1156860	1940290	24
6450_N LEAVITT_ST	205001010	0.00930737	1160420	1943000	24
6130_N_TALMAN_AVE	208002003	0.00258584	1157510	1940740	24
2711_W_FSTFRS_AVE	203004003	0.00258159	1156960	1946450	24

The #predictions generate a list of likely locations and published to an internal site used to route preventative baiting crews to bait likely locations.

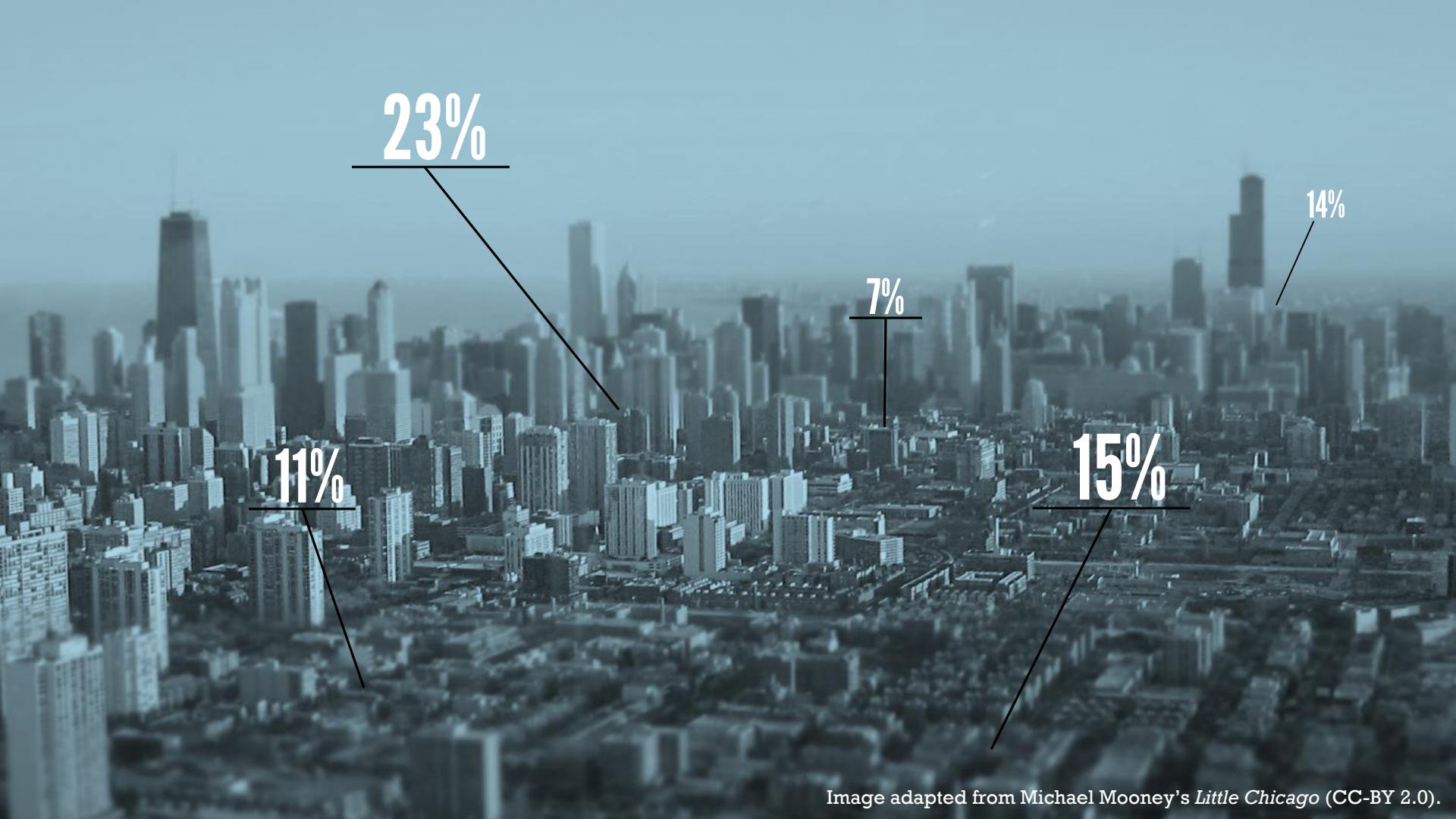


OPTIMIZING FOOD INSPECTIONS



OPTIMIZING FOOD INSPECTIONS





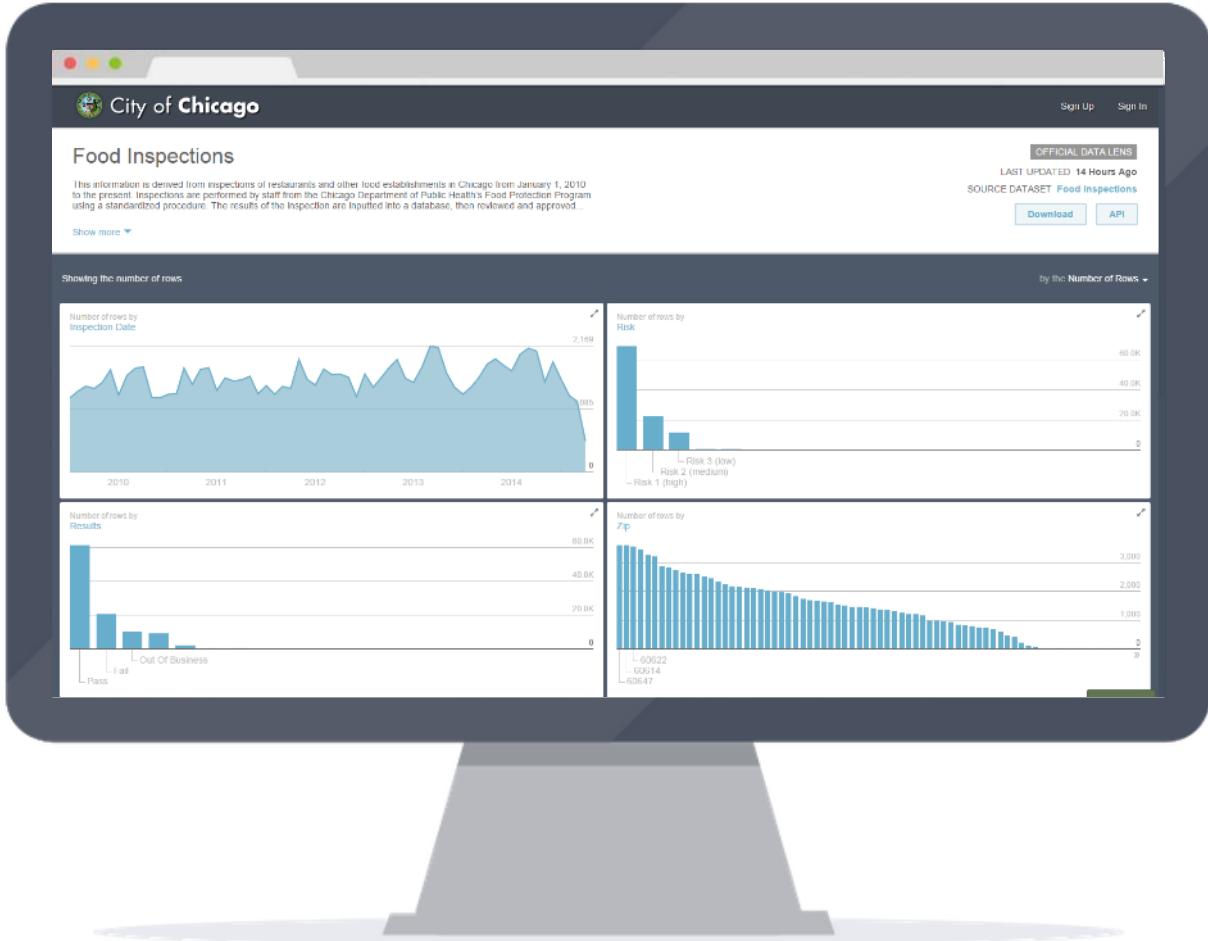
23%

11%

7%

15%

14%



USING OPEN DATA

Chicago leveraged the #opendata portal to share data with external researchers, leveraging the city's premiere method of sharing data and saving time on data-sharing agreements to create #predictions.

The model predicts the likelihood of a food establishment having a critical violation, a violation most likely to lead to food borne illnesses. Over a dozen #opendata sources were used to help define the model. Ultimately, ten different variables proved to create #predictions of critical violations.

Significant Predictors:

Restaurants with previous critical violations

Three-day average high temperature
CDPH risk level

Location of restaurant

Nearby garbage and sanitation complaints

Type of facility

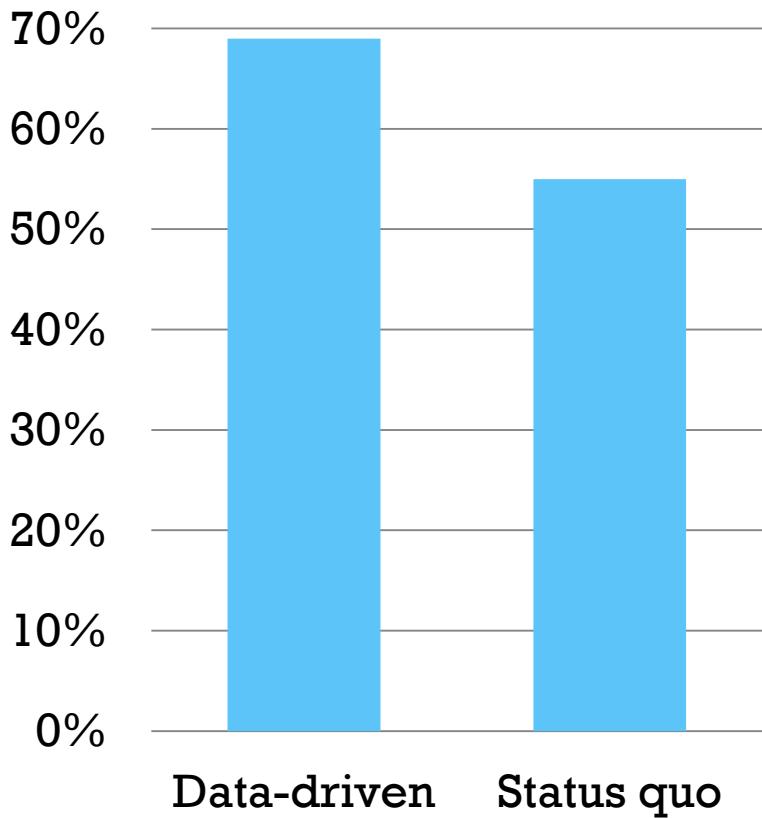
Nearby burglaries

Whether the establishment has a tobacco or has an incidental alcohol consumption license.

Length of time since last inspection.

Length of time the restaurant has been inspecting.

Critical violations



The #predictions revealed an opportunity to find deliver results faster. Within the first half of work, 69% of critical violations would have been found by inspectors using a data-driven approach. During the same period, only 55% of violations were found using the status quo method.

IMPROVEMENT

7 days

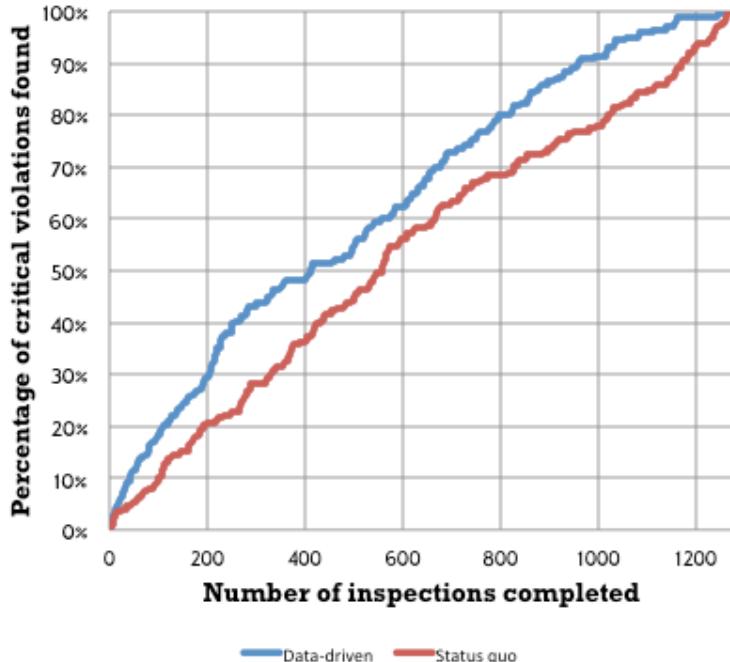
The food inspection model is able to deliver results faster.

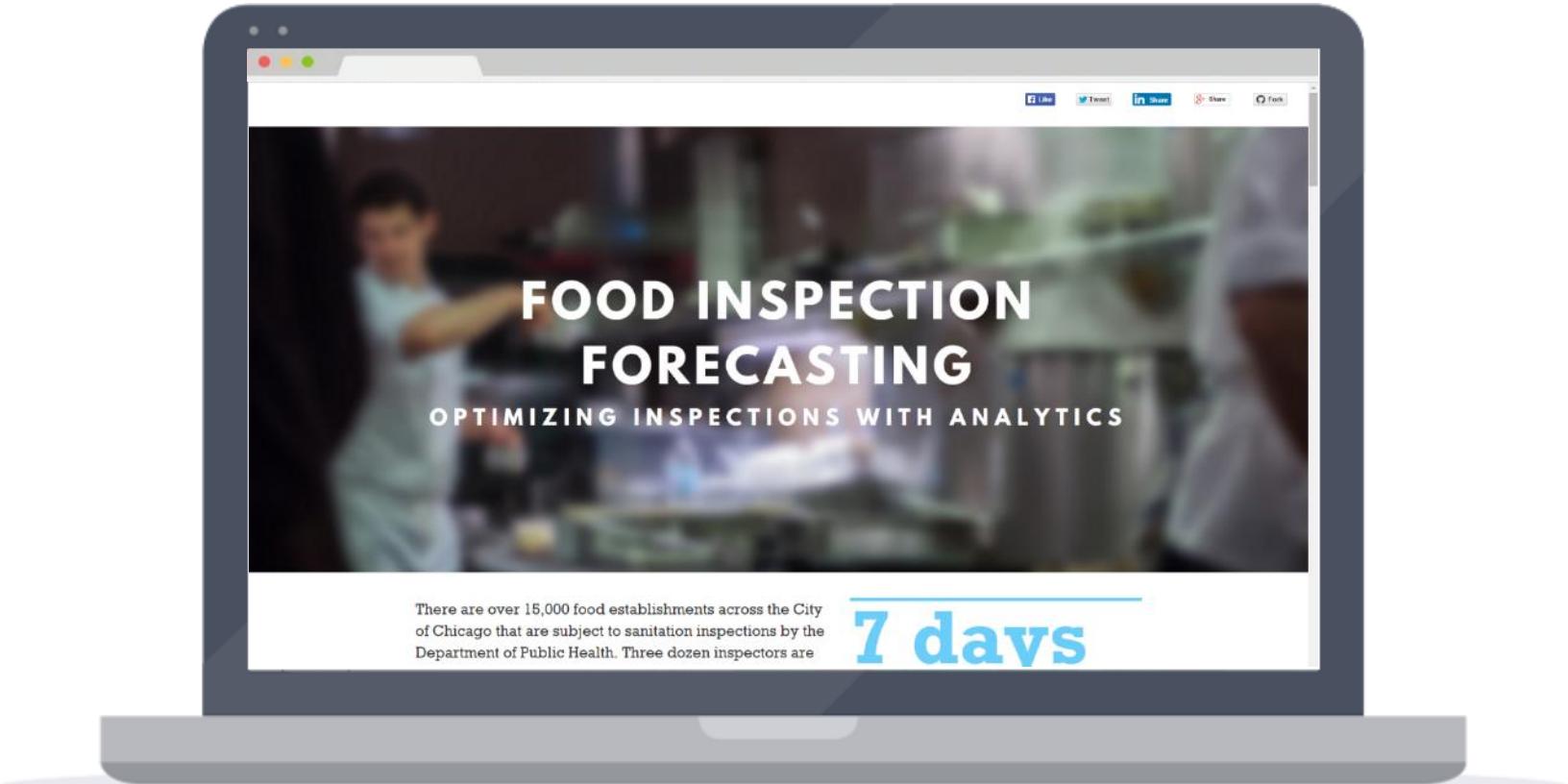
After comparing a data-driven approach versus the status quo, the rate of finding violations was **accelerated** by an average of 7.4 days in the 60 day pilot. That means the #predictions led to more violations would be found sooner by inspectors.

OPTIMIZING FOOD INSPECTIONS

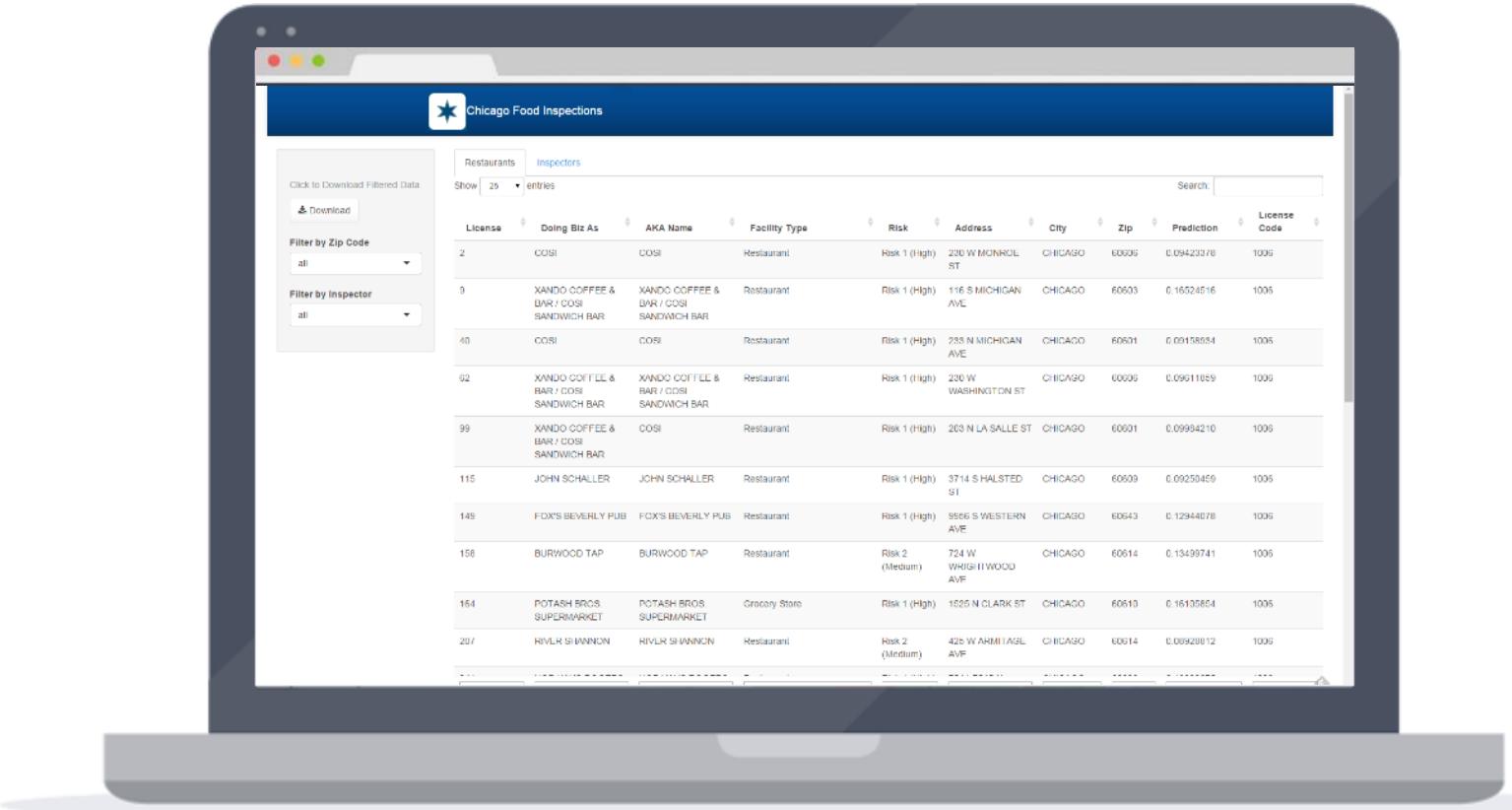
Discovering critical violations sooner rather than later reduces the risk of patrons becoming ill, which helps reduce medical expenses, lost time at work, and even a limited number of fatalities.

Impact





<http://chicago.github.io/food-inspections-evaluation/>



The data science team has built a website which lets CDPH prioritize inspections based on projected risk.

This screenshot shows the GitHub repository page for `Chicago / food-inspections-evaluation`. The repository is private. At the top, it displays 42 commits, 5 branches, 0 releases, and 1 contributor. The master branch is selected. The repository description states: "This repository contains the code to generate predictions of critical violations at food establishments in Chicago. It also contains the results of an evaluation of the effectiveness of those predictions. — Edit". The commit history shows several recent updates from `tomschenkjr`, including reading up-to-date weather data, updating datasets, fixing files, and adding compatibility notes. Below the commit history is a file listing for `README.md`. The main content area features a section titled "Food inspection forecasting evaluation" with a brief description of the repository's purpose. A sidebar on the right provides links to Code, Issues (4), Pull Requests (0), Wiki, Pulse, Graphs, and Settings. It also includes SSH clone URLs and options to Clone in Desktop or Download ZIP.

#OPEN SOURCE

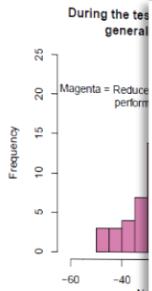
The analytical model will be released as an open source project on GitHub, allowing other cities to study or even adopt the model in their respective cities. No other city has released their analytic models before this release.

<http://github.com/Chicago/food-inspections-evaluation>

Likewise, weather is included since a single temperature series is only moving the entire predicted value of a food establishment one unit at a time. The objective model, which is the model presented in this report, [The instructions to run](#)

References

- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. *Regularization Paths for Generalized Linear Models via Coordinate Descent*. *The R Journal* 2 (1).
- Harris, Jerome K., East Man Department Use of Social Media Mortality. *Worldly Report* 63 (3).
- Simon, Noah, Jerome Friedman, and Trevor Hastie. 2010. *Cox's Proportional Hazards Model: The R Function chfitter*. <http://www.jstatsoft.org/v032/i03>.
- Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Springer.



We conducted a t-test to measure whether the mean number of critical violations was zero. Namely, the null hypothesis is that the mean number of critical violations is zero. The test ($\sigma = 25.0$, $df = 257$) resulted in a p-value of 0.0001, which is significantly less than 0.05.

Below, Gini curves show the relative first day cumulative data points model versus the first week between Sep 2013 and Sep 2014. The cumulative approach until the final day of the year is shown.

Forecasting restaurants with critical violations in Chicago

Gene Leynes (City of Chicago), Aakash Solanki (City of Chicago), Tim Schenk, Jr. (City of Chicago)
February 16, 2015

The Chicago Department of Public Health (CDPH) inspects more than 15,000 restaurants with fewer than three dozen inspectors over the course of the year. This paper describes a predictive model designed to identify the presence of a critical violation in a particular food establishment. This model can help prioritize inspections by likelihood in order to identify the riskiest restaurants earlier, thereby reducing the length of exposure of risky restaurants to patrons. Critical violations were identified approximately 7.44 days earlier over a 90-day period compared to current operations in the out-of-sample test.

1 Introduction

In 2014 the Chicago Department of Public Health inspected performed over 20,000 inspections at nearly 13,000 food establishments across Chicago with fewer than three dozen inspectors. The majority of those food inspections were routine inspections that didn't uncover serious problems, but some of those inspections uncovered issues that affect the health and safety of the patrons who visit these establishments. Traditionally, prioritizing inspections is a largely manual task that relies on a combination of administrative processes and personal expertise.

The model set forth in this paper can help with the prioritization of scheduled, ongoing time and money as well as making the city's food safer. The model utilizes several data sources and through advanced modeling techniques the model provides additional insight into an establishment's current actual risk based on real-time data.

This paper is organized as follows: Section 1 provides an introduction and background to describe the current process and scope of the problem. Section 2 describes data that has been collected by the research team for this project, and how that data was used to build the model. Section 3 discusses the model results. Section 4 discusses the model results in comparison to the empirical inspection data. Finally, the Summary section concludes with a brief summary of results and information regarding the ongoing project.

Ultimately, we find that a data-driven model can help inspectors discover critical violations earlier than the current "Business As Usual" (BAU) process. On average, critical violations would have been discovered 7.44 days earlier over the two-month test period. The first half of the experiment yielded 25.0% higher successful inspections. Beginning in 2015, CDPH has begun to use this empirical model to prioritize routine inspections. Inspectors will still conduct visual and smell portions; however, these restaurants with the highest risk of critical violations will be prioritized.

It is worth noting that this research is an open source project. The source code of the statistical model is available online at the City of Chicago food inspection project page. The statistical modeling was completed using the open source statistical software R, and all the necessary data to replicate these results is available online. This paper was generated using knitr, which allows others to view the underlying calculations to generate the summaries, tables, and diagrams in this document. This document is available in the same aforementioned repository.

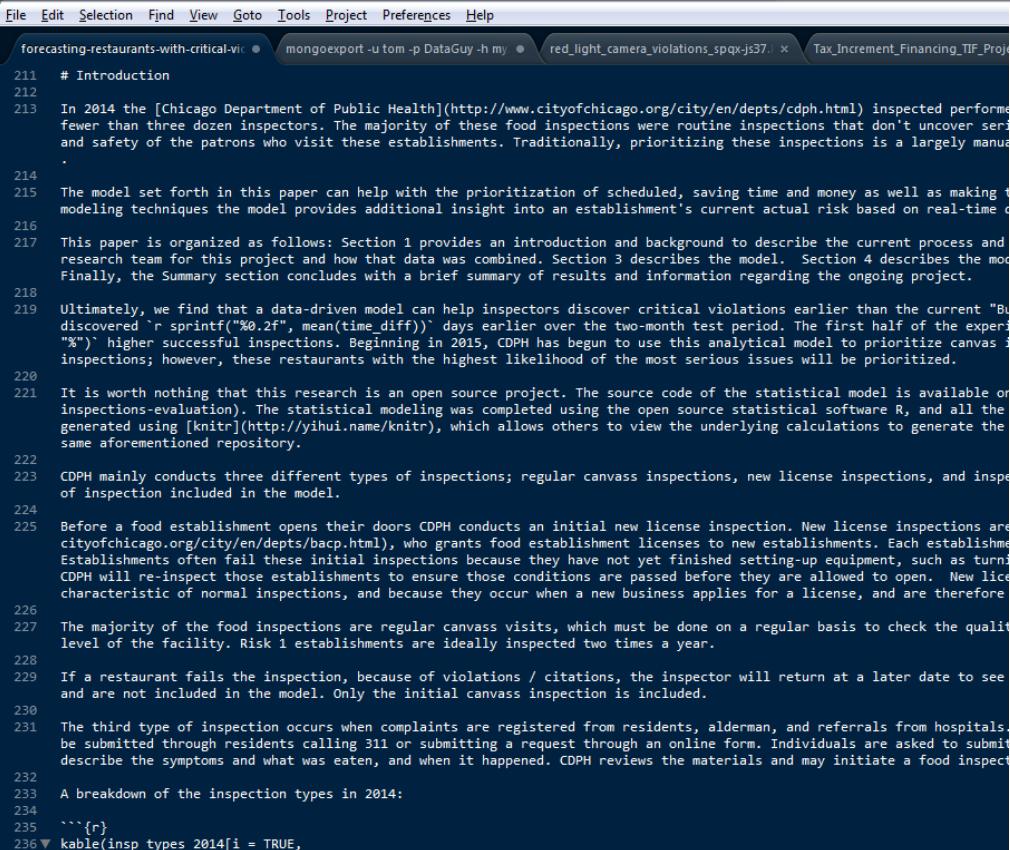
CDPH mainly conducts three different types of inspections: regular routine inspections, new license inspections, and inspections in response to a complaint. Currently regular routine visits are the only type of inspection included in the model.

Technical Documentation

The project was released using an academic-quality technical paper instructing others on the the variables and statistical methodology used in the project. In addition to source code, the paper will help researchers adopt this approach.

Reproducible Research

The technical paper was written as a highly-reproducible “knitr” document, allowing other researchers to understand how summary numbers were calculated. Each statement in the project can be traced to an original source.



A screenshot of a code editor window showing an R script. The window has a menu bar with File, Edit, Selection, Find, View, Goto, Tools, Project, Preferences, and Help. Below the menu is a tab bar with several tabs: "forecasting-restaurants-with-critical-vi...", "mongoexport -u tom -p DataGuy -h my...", "red_light_camera_violations_spqx-js37...", and "Tax_Increment_Financing_TIF_Proj...". The main area contains R code with line numbers from 211 to 236. The code discusses food inspections by the Chicago Department of Public Health (CDPH) in 2014, the use of a statistical model to prioritize inspections, and the types of inspections conducted by CDPH. It also mentions the use of knitr to generate the report.

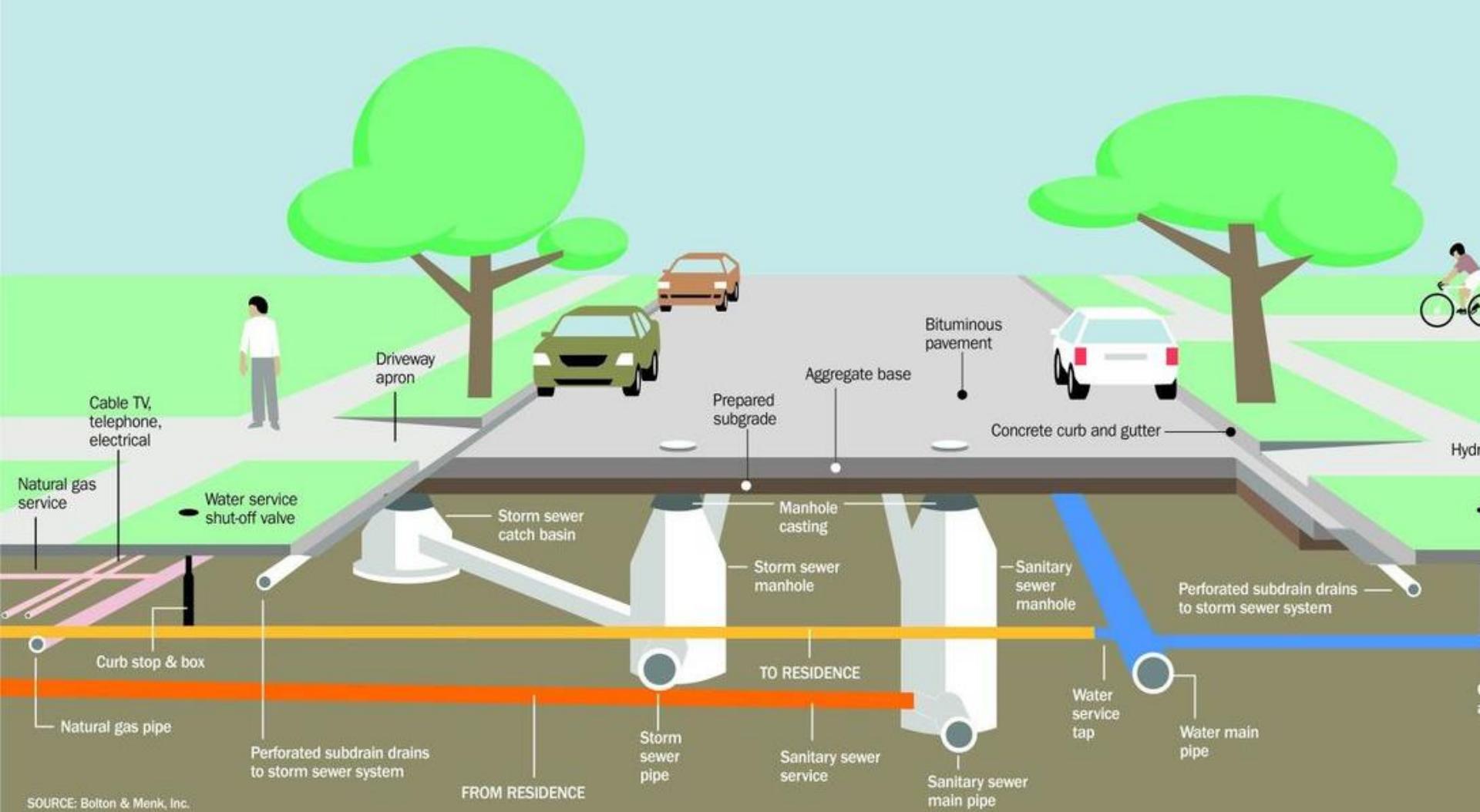
```
File Edit Selection Find View Goto Tools Project Preferences Help

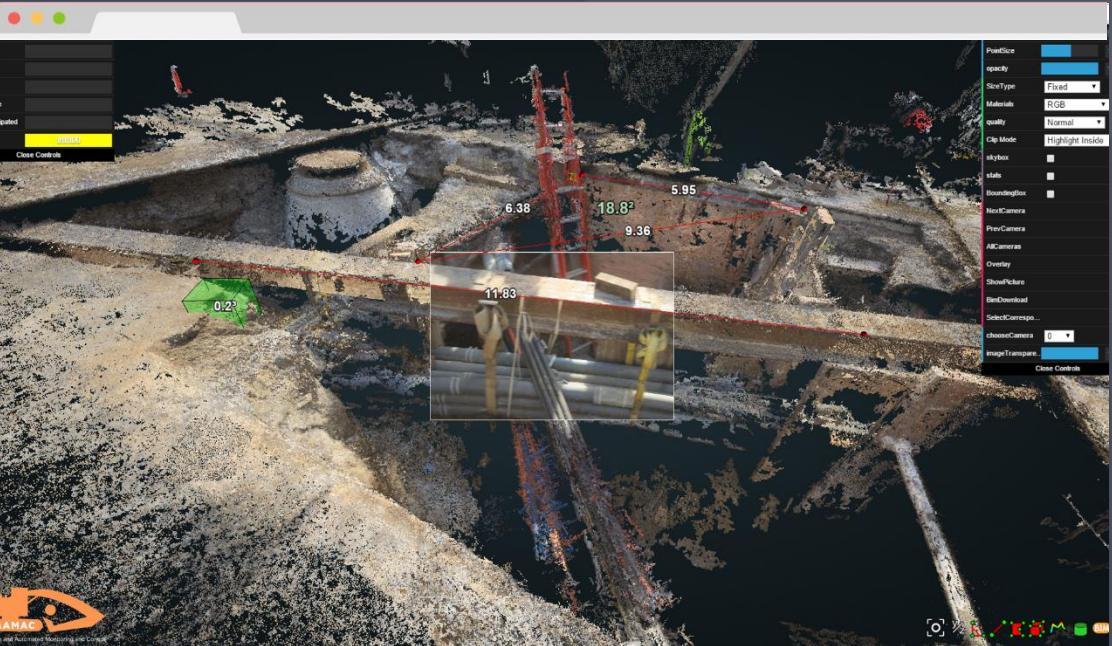
forecasting-restaurants-with-critical-vi... mongoexport -u tom -p DataGuy -h my... red_light_camera_violations_spqx-js37... Tax_Increment_Financing_TIF_Proj...

211 # Introduction
212
213 In 2014 the [Chicago Department of Public Health](http://www.cityofchicago.org/city/en/depts/cdph.html) inspected fewer than three dozen inspectors. The majority of these food inspections were routine inspections that don't uncover serious and safety of the patrons who visit these establishments. Traditionally, prioritizing these inspections is a largely manual process.
214
215 The model set forth in this paper can help with the prioritization of scheduled, saving time and money as well as making the modeling techniques the model provides additional insight into an establishment's current actual risk based on real-time data.
216
217 This paper is organized as follows: Section 1 provides an introduction and background to describe the current process and research team for this project and how that data was combined. Section 3 describes the model. Section 4 describes the model's results.
218
219 Ultimately, we find that a data-driven model can help inspectors discover critical violations earlier than the current "Business as Usual" approach. For example, the model discovered `r sprintf("%0.2f", mean(time_diff))` days earlier over the two-month test period. The first half of the experiments resulted in `r sprintf("%d", nrow(experiments))` higher successful inspections. Beginning in 2015, CDPH has begun to use this analytical model to prioritize canvas inspections; however, these restaurants with the highest likelihood of the most serious issues will be prioritized.
220
221 It is worth nothing that this research is an open source project. The source code of the statistical model is available on GitHub at https://github.com/yihui/knitr. The statistical modeling was completed using the open source statistical software R, and all the generated reports are generated using [knitr](http://yihui.name/knitr), which allows others to view the underlying calculations to generate the same aforementioned repository.
222
223 CDPH mainly conducts three different types of inspections: regular canvass inspections, new license inspections, and inspection of inspection included in the model.
224
225 Before a food establishment opens their doors CDPH conducts an initial new license inspection. New license inspections are conducted by the City of Chicago (http://www.cityofchicago.org/city/en/depts/bacp.html), who grants food establishment licenses to new establishments. Each establishment must pass this inspection before they are allowed to open. Establishments often fail these initial inspections because they have not yet finished setting-up equipment, such as turning on refrigerators. CDPH will re-inspect those establishments to ensure those conditions are passed before they are allowed to open. New license inspections are characteristic of normal inspections, and because they occur when a new business applies for a license, and are therefore called "new license inspections".
226
227 The majority of the food inspections are regular canvass visits, which must be done on a regular basis to check the quality level of the facility. Risk 1 establishments are ideally inspected two times a year.
228
229 If a restaurant fails the inspection, because of violations / citations, the inspector will return at a later date to see if the violations have been corrected. These follow-up inspections are not included in the model. Only the initial canvass inspection is included.
230
231 The third type of inspection occurs when complaints are registered from residents, alderman, and referrals from hospitals. These complaints are submitted through residents calling 311 or submitting a request through an online form. Individuals are asked to submit a detailed report describing the symptoms and what was eaten, and when it happened. CDPH reviews the materials and may initiate a food inspection.
232
233 A breakdown of the inspection types in 2014:
234
235 ````{r}
236 kable(insp_types_2014[i = TRUE,
```



**UNDERGROUND
INFRASTRUCTURE IS HIT ON
AVERAGE EVERY 60 SECONDS.
THE TOTAL COST TO THE
NATIONAL ECONOMY IS
ESTIMATED TO BE \$1.6 BILLION**

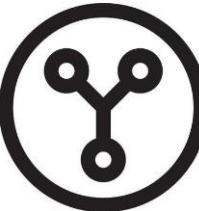




Underground Map

Using off-the-shelf DSLR cameras, photos are stitched together to create a 3-D model of the city's underground infrastructure. City Digital, City of Chicago and a consortium of partners are piloting the tech.

THANK YOU



Tom Schenk Jr.

Chief Data Officer

City of Chicago

Contact Info:

@ChicagoCDO

tom.schenk@cityofchicago.org

Websites:

Data @ data.cityofchicago.org

News @ digital.cityofchicago.org

Maps @ opengrid.io

Strategy @ techplan.cityofchicago.org

Slides @ speakerdeck.com/tomschenkjr