BlueConduit

Using AI to reduce uncertainty of lead service line replacement

Dr. Eric Schwartz, Assistant Professor, Ross School of Business, Univ. of Michigan, Co-founder of BlueConduit
Dr. Jacob Abernethy, Assistant Professor, College of Computing, Georgia Tech, Co-founder of BlueConduit
Ian Robinson, Managing Director, BlueConduit
Brigadier General (ret.) Michael McDaniel, Director of Government and Client Services
Jared Webb, Chief Data Scientist, BlueConduit

Jake Abernethy, Ph.D
Co-Founder.

Professor of Machine Learning (ML) at the College of Computing at Georgia Tech. 15 years in machine learning. Applied data science to help Flint since 2016.

Eric Schwartz, Ph.D..
Co-Founder

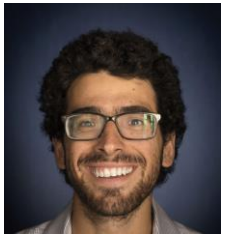
Professor of marketing at the Ross School of Business at the University of Michigan. Worked in Flint since 2016

B.G. (ret.) Michael McDaniel
Head of Government and Client Services

Coordinated lead service line replacement for Flint's FAST Start program. Dean of WMU-Cooley's Homeland and National Security Law program

Jared Webb
Chief Data Scientist

MS in Math from Brigham Young University. Applied data science to help Flint since 2016.



Ian Robinson
Managing Director

MBA/MS from University of Michigan. Experience in growing social enterprises. Returned Peace Corps Volunteer

Solution: We Bring Clarity to the Unknown

*Using **cutting-edge artificial intelligence** methods, we characterize and resolve uncertainties around the material of public and private* service lines.

Actionable Data

- ✔ *Prioritize high-likelihood lead service lines*
- ✔ *Maximize public health benefits*
- ✔ *Optimize costs*

Accurate LSL Inventories

- ✔ *Enable cities to budget for SL replacement*
- ✔ *Guide current and future asset management strategies*
- ✔ *Improve program cost and management*

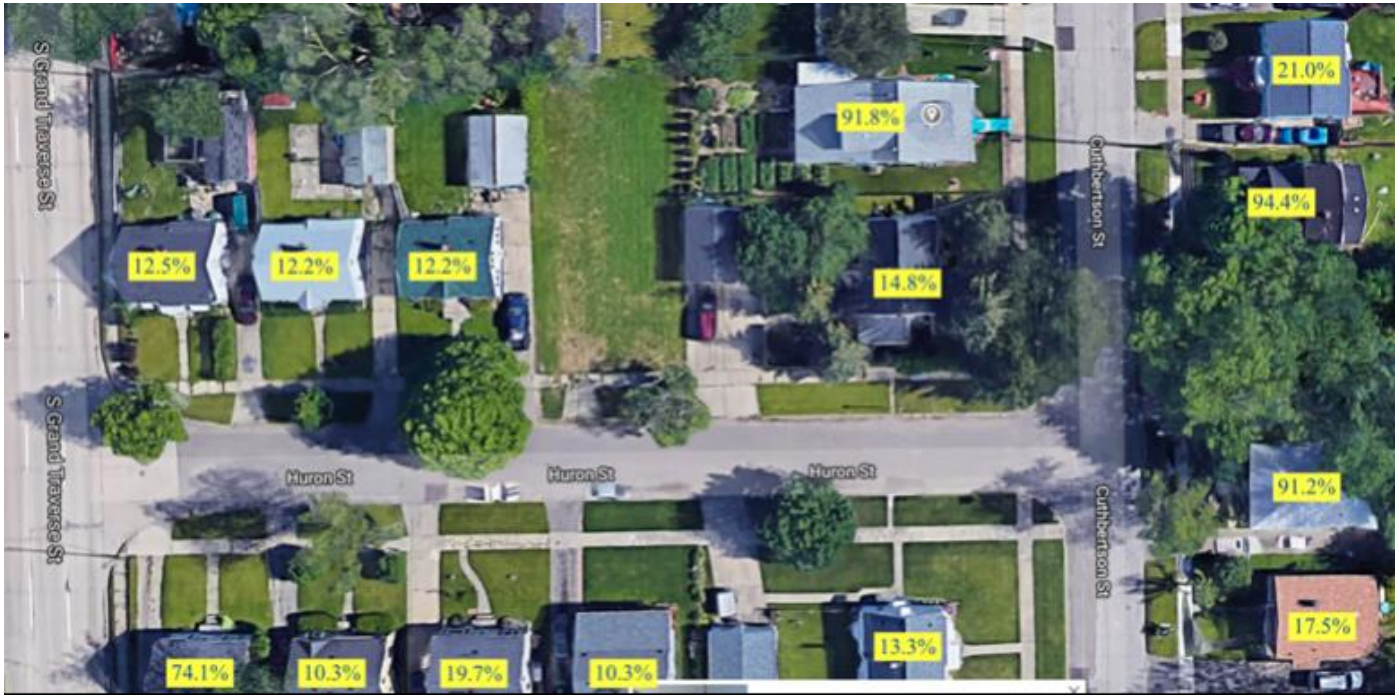
Trust & Communication

- ✔ *Provide foundation for local education*
- ✔ *Develop tools for community engagement*
- ✔ *Support regulatory compliance*

Predicted chance of each lead/galvanized service line for these Flint homes, as of May 2018, according to predictive model

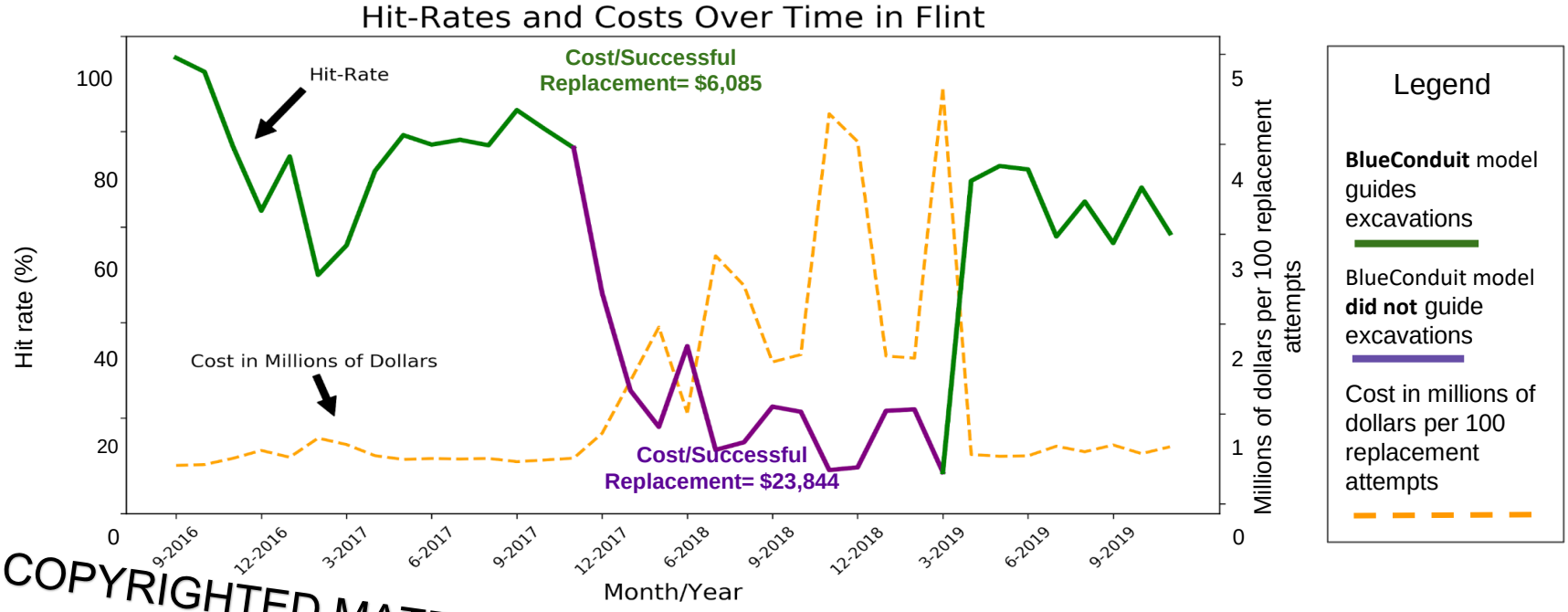
Note that homes with 91% likelihood are next to and across the street from homes with 14% and 12% probability

Based on standard statistical methods and a published application for machine learning



Higher Hit Rate Reduces Costs

Would have saved 20-30% by avoiding unnecessary excavations out of $100MM budget



Cities don't know number and location of lead service lines

American Water Works Association

Replacing all lead water pipes could cost $30 billion

EPA United States Environmental Protection Agency

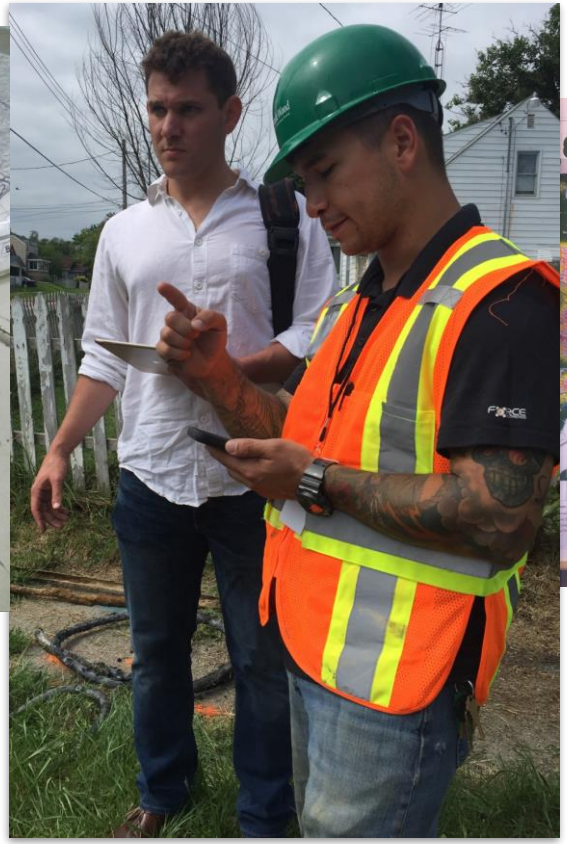
Lead and Copper Rule Long-Term Revisions

EDF ENVIRONMENTAL DEFENSE FUND
Finding the ways that work

"The current LCR disproportionately impacts low-income and minority populations"

"Every lead service line replaced yields an estimated $22,000 in reduced cardiovascular disease deaths"

The Work We Did In Flint



Three-Phase Approach

	1 Preliminary Estimates	2 Home-by-Home Predictive Recommendations	3 Continuous Improvement
INPUTS	Receive existing data	Verify material at recommended sites	Real-time SL data from replacements, field work
OUTPUTS	**Recommended inspection list** Region-level preliminary estimates	**Prioritized replacement list** Water filter and public health communication support	**Updated prioritized list** Continued AI refinement Optional: Community-facing map portal

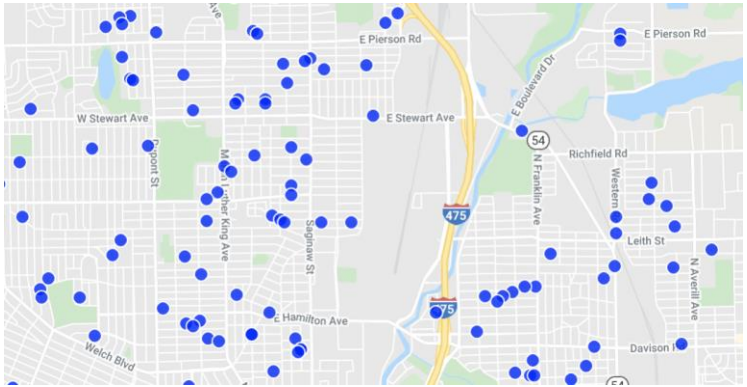
1 Preliminary Estimates

INPUTS

- ❑ Recent SL material confirmation
- ❑ Historical SL records
- ❑ Parcel attribute data

OUTPUTS

- ✔ **Recommended Inspection List**
- ✔ Representative data increases accuracy
- ✔ Region-level preliminary estimates
- ✔ Budget estimates



Recommended Inspections

Region	% of city	Copper	Lead	Galvanized	Plastic
Ward 1	21%	70	20	10	5
Ward 2	28%	80	50	5	0
Ward 3	16%	40	30	5	4
Ward 4	14%	20	40	10	0
Ward 5	21%	90	5	3	3

Region-level preliminary estimates

* Recent SL material confirmation sites (e.g., water main break, service line leaks, meter replacements) do not reflect the whole system.

Inspections to Gather Data are the Essential First Step

Before BlueConduit
Historical records suggested **10-20%** of SL were lead

The first 170 attempted replacements found lead **96%** of the time

With BlueConduit
2016: BlueConduit-recommended inspections at a representative set indicated **49%** SLs had lead in the inventory

2019: The true percentage of Flint homes that contained lead/galvanized (after 20,000 excavations) was verified **51%**

1 What data do we use?

SL Verified (Private)	SL Verified (Public)	Parcel ID	Address	Year Built	Latitude	Longitude	Home value	SL Historical Record (Private)	SL Historical Record (Public)	Water Test Results (ppb)
Galvanized	Copper	401446	6201 MAPLE	1925	43.00802	-83.6365	7800	Copper	Copper	0
?	?	401215	1234 MAIN	1957	43.07282	-83.7219	19300	Copper	-	2
?	?	410545	567 WALNUT	1921	43.02797	-83.7171	10900	Copper	Lead	8
Copper	Copper	401438	1486 OAK	-	43.02454	-83.6655	4000	Copper	Copper	-
Lead	Lead	462645	2014 ELM	1970	43.05868	-83.725	17700	-	-	-

Information inputs:

- Municipalities providing available parcel-level data for each service point
- BlueConduit obtains external data sources to improve models

Model adapts to areas with many or minimal historical records

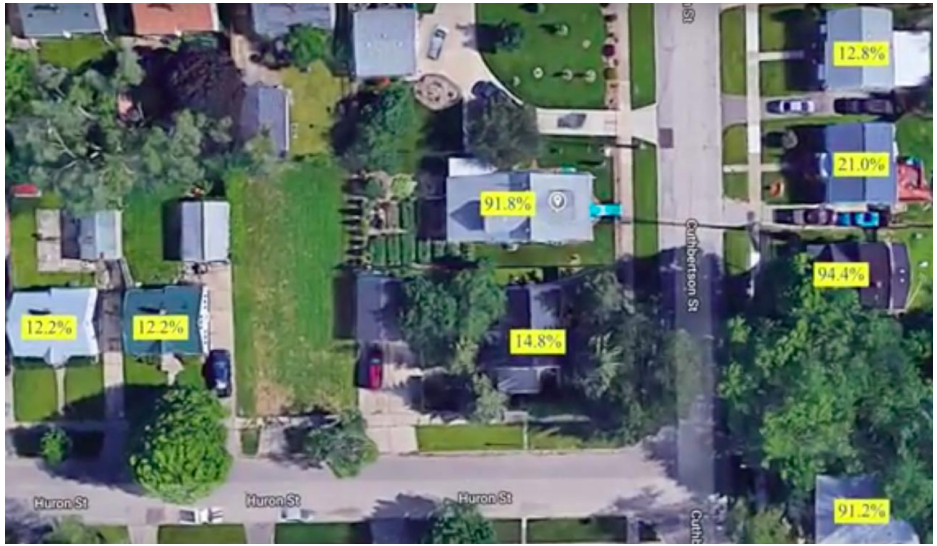
2 Home-by-Home Predictive Recommendations

INPUTS

- ❑ Send crews to sites on **Recommended Inspection List** to verify SL material

OUTPUTS

- ✔ **Prioritized Replacement List**
  - Home-by-Home probabilities
  - Map layer and database
  - Organized at home-, block-, or main-level
- ✔ System-wide home level inventory
- ✔ Support public health and compliance
  - Target filters, comms, testing



Probabilities generated by statistical machine learning using approach from Abernethy et al. (2018)

Prioritization Strategy for Efficiency and Equity

Organize by individual service probabilities:

Home Address	Priority	Probability
6275 Summit Pl	High	80.40%
2277 Commonwealth Ave	Low	2.10%
8764 Liverpool St	Medium	41%
28855 Rockeldge	Low	3.42%
1329 S University	High	94.52%

Home Address	Priority	Probability
1329 S University	High	94.52%
14 Harvard Rd	High	91.23%
637 Baltic Ave	High	90.78%
1043 Fish Hatchery Rd.	High	89.45%
6275 Summit Pl	High	80.40%

Aggregate by water mains:

Water Main No.	Priority	Num SLs
WS 1020	Low	0.044
WS 1244	Medium	1.2
WS 1234	High	6.8

Water Main No.	Priority	Num SLs
WS 1234	High	6.8
WS 2646	High	6.2
WS 1923	High	5.8

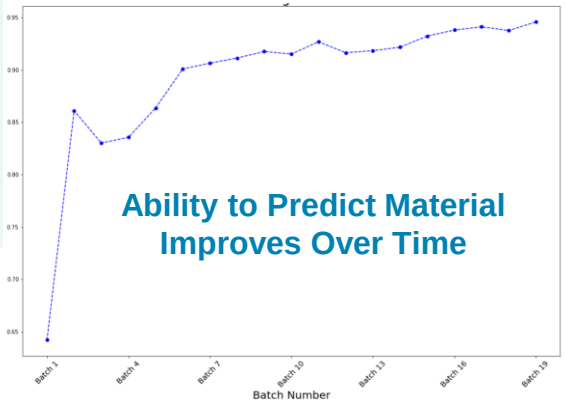
3 Continuous Improvement

INPUTS

- ❑ Share real-time SL data from prioritized replacements and other field work

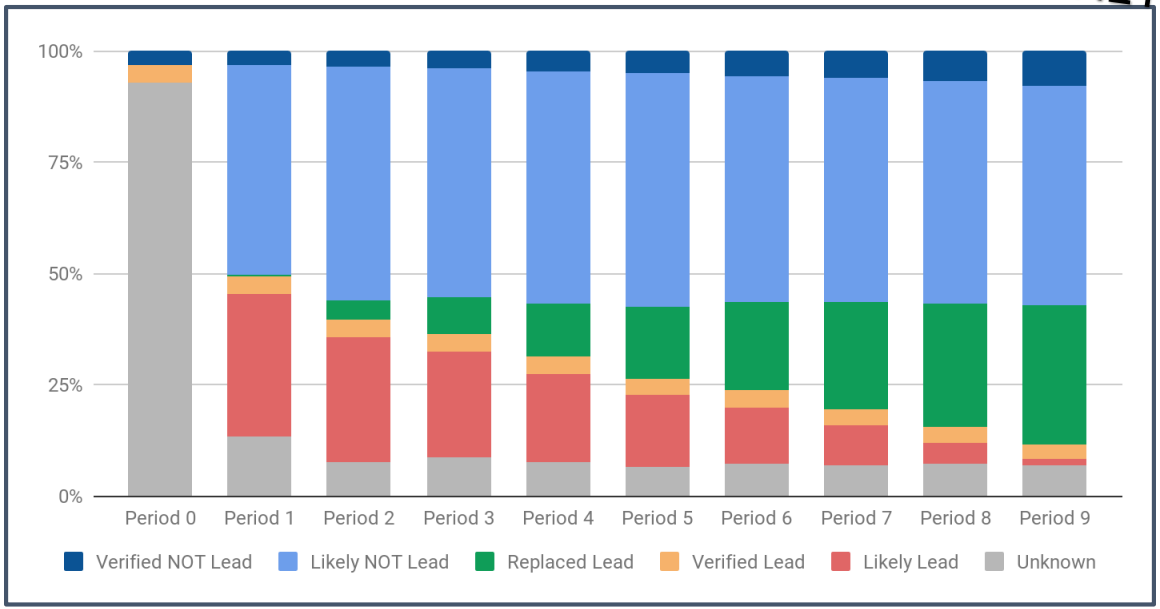
OUTPUTS

- ✔ **Updated Prioritized List**
  - Continued AI model refinement
- ✔ Optional: Community-facing map portal



Optional: Community-facing map portal

3 Digging Where the Lead Is



Continuous improvement *leads to* ***saved time and money*** *from excavating only the pipes likely to need replacement, and on rapidly improved to* ***public health outcomes****.*

Making this solution accessible

We want to help communities do this right now so they get ahead of this problem

Working with local and national foundations to subsidize cost to communities

We have already secured funding from The Rockefeller Foundation to work with 4 cities in 2020

- EPA Environmental Justice grant with Toledo
- Benton Harbor

Funding is available to communities that meet environmental justice and socioeconomic criteria

Input Data

1
- Recent SL material confirmation
- Historical SL records
- Parcel attribute data

2
- New SL material verification

3
- Real-time data from prioritized replacements and other field work

Output Data

1
- Recommended Inspection List
- Region-level preliminary estimates

2
- Prioritized Replacement List
- System-wide home level inventory

3
- Continued AI refinement

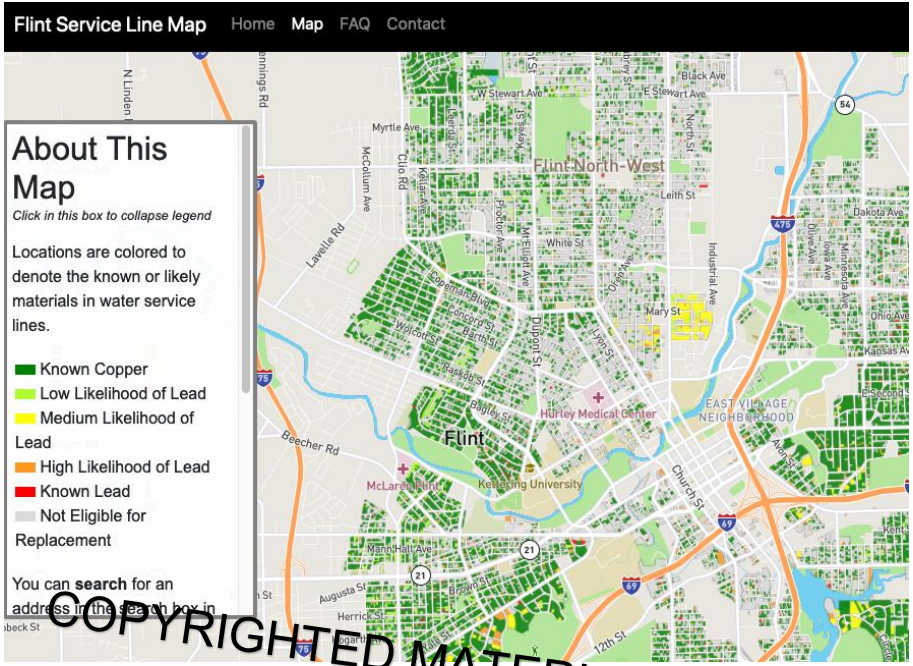
BlueConduit

Thank You

ian@blueconduit.com

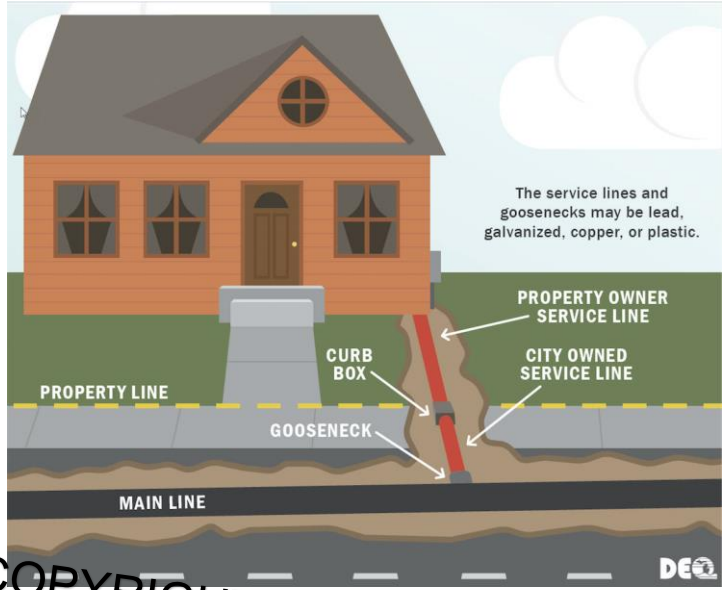
Publicly accessible map portal for residents to search

- Tool for transparency
- Based on city data and predictive algorithm
- Collaboration with NRDC public health experts

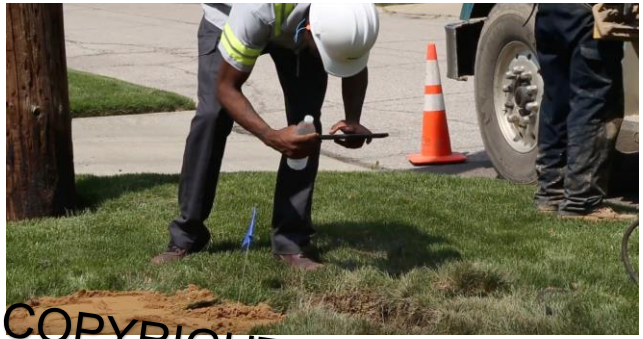


Barriers to efficient inventory and replacement

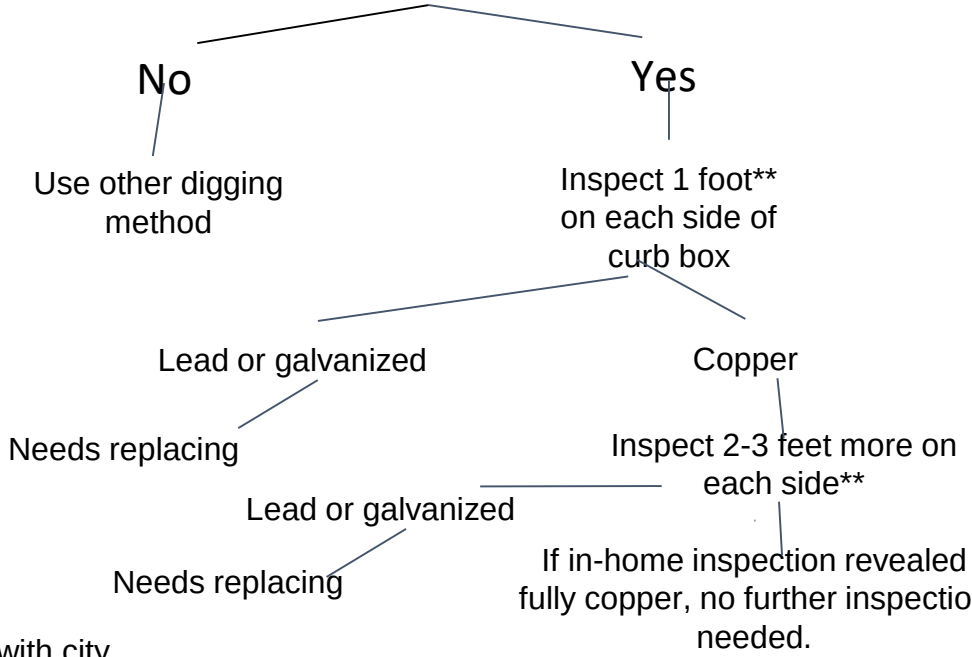
- Unreliable, outdated, and missing data about LSL locations
- Multiple segments do not made of same material
- Expensive to verify pipe material



Information Gathering via Inspections: Using Hydrovac



Example decision tree



** Exact numbers determined with city

BlueConduit

Using data from inspections

- The proportion of lead/galvanized service lines in the sample should be close to the true proportion of lead/galvanized service lines in the population.
- Initial representative inspection set allowed Flint to budget and plan
- Give more accurate estimates for annual SL replacement rates
- Key input to developing home-by-home probabilities

Gathering More Data with a Representative Initial Inspection Set

- Initially, inspections should be done on a **Representative Inspection Set**
  1. Existing data is biased and not representative of the entire system
     a. E.g., Cities may have SL materials verified ONLY where water mains have been replaced due to breaks. But the blocks and areas where mains have broken may be more likely to have certain kind of SL materials.
  2. Planning and budgeting requires system-wide estimates
  3. Selecting one or a few separate neighborhoods may not generalize.
     a. Could mischaracterize other neighborhoods at greater risk.
  4. Improves predictive accuracy of statistical models extrapolating materials for homes with unknown SL materials.

Representative Initial Inspection Set

- Inspections are more efficient
  - If all homes could be inspected immediately, problem scope would be defined. But they cannot be…
  - Cheaper and faster data gathering tool.
  - If concerns over accuracy (e.g., hydrovac) and “false negatives” (e.g., lead was not uncovered, when there was truly lead), then
    - Build in guidelines and redundancies
    - Performance-based contracts with bonuses/penalties

Integrating data points

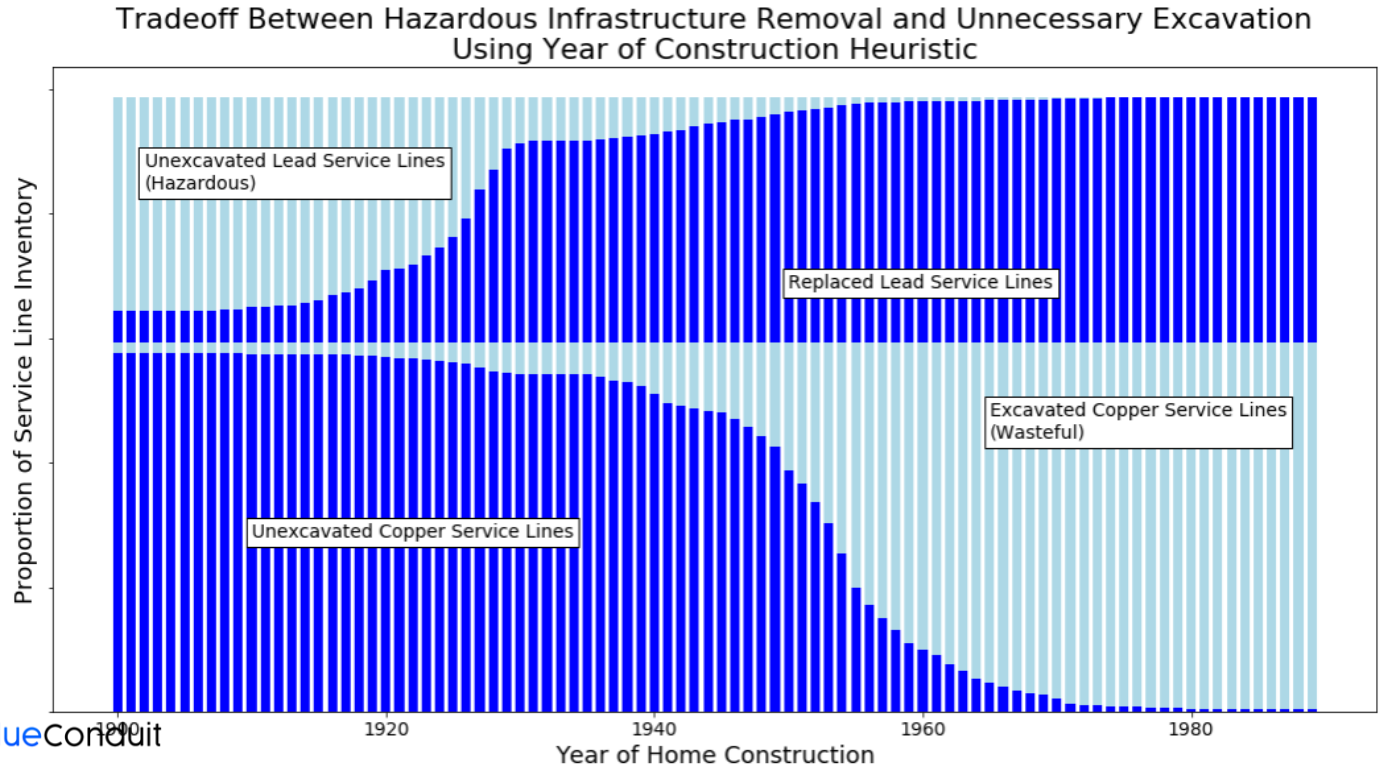
Relying on water test results in Flint would have resulted in 2761 unnecessary excavations and overlooked 702 homes with lead service lines.

	Water test > 0 ppb	**Water test = 0 ppb**
Lead Service Line	2935	702
Non-lead Service Line	2761	2930

A machine learning model integrates all of this data in making predictions.

Integrating data points

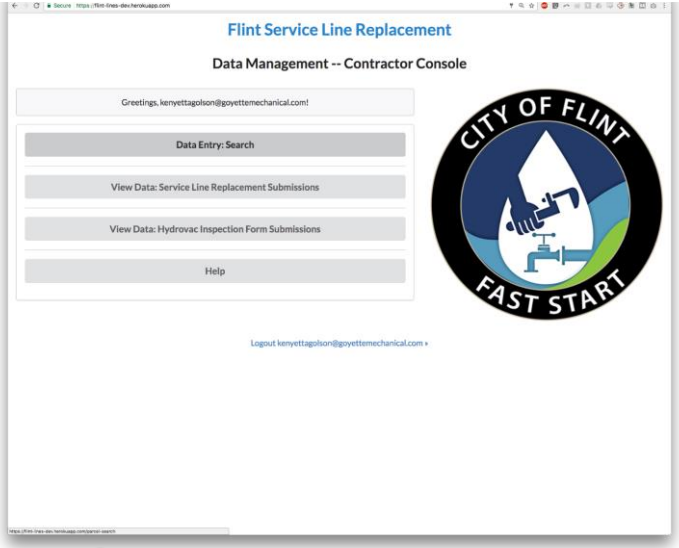
- Examining accuracy of “year built” labels and Verified Materials via inspection or replacement



Note: Year built can be misleading, especially in the case of teardowns

Data Collection Webapp V2.0: Demo (images linked to web)

Data collection app built for Flint contractors to input necessary information in required format to populate database



Sample size in different-sized communities

Sample size required to achieve a Margin of Error of 5% and Confidence Interval of 95%

	Total number of service lines			
	1000	**5000**	**10000**	**50000**
Inspections in initial set	278	357	370	382

"With a Representative initial set of this size, the observed percent of SL needing replacement plus/minus 5% will include the true proportion in 95% or 19 out of 20 times." So "if you observe 40% then the estimated range of 35% to 45% will be right in 19 times out of 20!

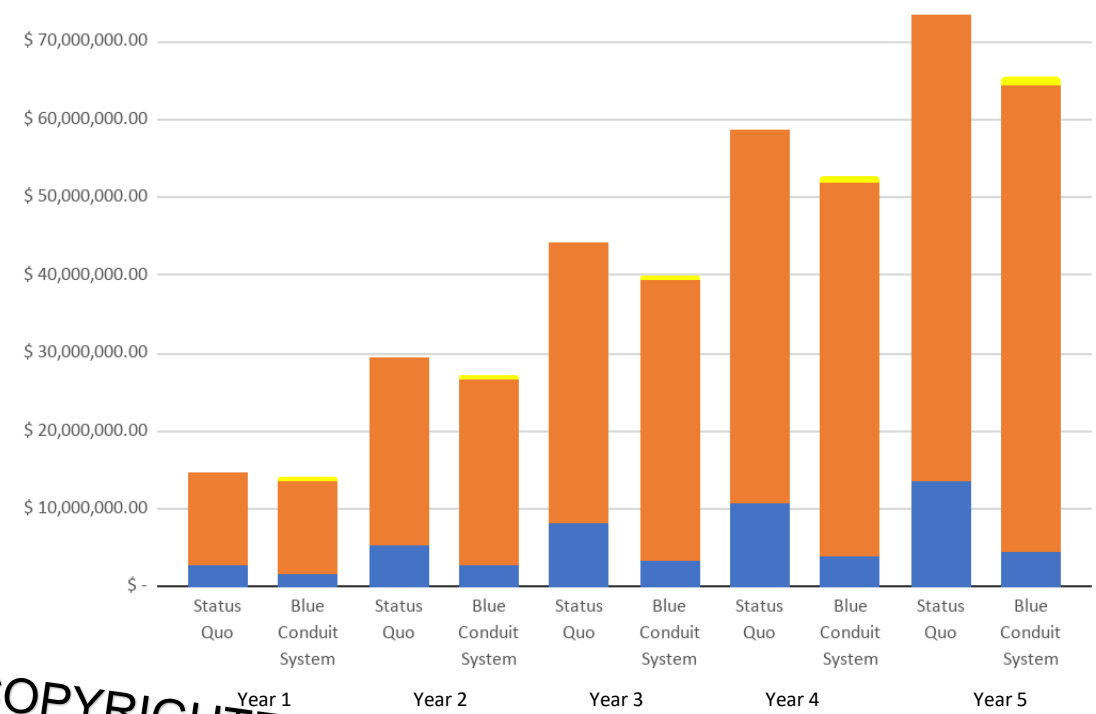
Sample size required to achieve a Margin of Error of 5% and Confidence Interval of 99%

	Total number of service lines			
	1000	**5000**	**10000**	**50000**
Inspections in initial set	400	586	623	655

Key assumptions
1. You try to estimate one number: the proportion of SL needing replacement system wise.
2. Conservative guess at true proportion to be 50%.

More inspection sites required when considering multiple areas

Cumulative Costs Over Five Years



Cost savings	$8,300,000
Avoided excavations	3000

Successful Replacements per year	2000
Status quo hit rate	55%
BlueConduit method hit rate in year 1	0.75
BlueConduit method hit rate in year 2	0.8
BlueConduit method hit rate in year 3-5	0.9
Year 1 Inspections	1000
Year 2-5 Inspections	200
Cost of Excavation	$3,000
Cost of Replacement	$3,000
Cost of Hydrovac/ Inspection	$500

Design Contracts and RFP's to Optimize Efficiency

Important lesson learned from Flint: having separate contractors for inspections and replacements reduces incentive misalignment

Replacement Contractors

Send contractors with the intention of **replacing Lead/Galvanized SL** to areas with **highest likelihood**
Goal: replace as many lines as possible
KPI's/ Incentives:

- **Effective cost of successful service line replacement ($**) = Total cost of all replacement attempts / # of successful replacements
- **Hit rate (%)**= Success/ Attempts

Inspection Contractors

Send contractors with purpose of **inspections SL** to all homes **(target with highest uncertainty)**
Goal: Collect as much accurate SL material labels
KPI's/ Incentives:

- **# of confirmed lines**