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
Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Replacing all lead water pipes could cost $30 billion

Lead and Copper Rule Long-Term Revisions

“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
   - INPUTS: Receive existing data
   - OUTPUTS: Recommended inspection list, Region-level preliminary estimates

2. Home-by-Home Predictive Recommendations
   - INPUTS: Verify material at recommended sites
   - OUTPUTS: Prioritized replacement list, Water filter and public health communication support

3. Continuous Improvement
   - INPUTS: Real-time SL data from replacements, field work
   - OUTPUTS: Updated prioritized list, Continued AI refinement, Optional: Community-facing map portal
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**

<table>
<thead>
<tr>
<th>Region</th>
<th>% of city</th>
<th>Copper</th>
<th>Lead</th>
<th>Galvanized</th>
<th>Plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ward 1</td>
<td>21%</td>
<td>70</td>
<td>20</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Ward 2</td>
<td>28%</td>
<td>80</td>
<td>50</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Ward 3</td>
<td>16%</td>
<td>40</td>
<td>30</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Ward 4</td>
<td>14%</td>
<td>20</td>
<td>40</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Ward 5</td>
<td>21%</td>
<td>90</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

* Recent SL material confirmation sites (e.g., water main break, service line leaks, meter replacements) do not reflect the whole system.
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%**
## What data do we use?

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

### 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
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:

<table>
<thead>
<tr>
<th>Home Address</th>
<th>Priority</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>6275 Summit Pl</td>
<td>High</td>
<td>80.40%</td>
</tr>
<tr>
<td>2277 Commonwealth Ave</td>
<td>Low</td>
<td>2.10%</td>
</tr>
<tr>
<td>8764 Liverpool St</td>
<td>Medium</td>
<td>41%</td>
</tr>
<tr>
<td>28855 Rockledge</td>
<td>Low</td>
<td>3.42%</td>
</tr>
<tr>
<td>1329 S University</td>
<td>High</td>
<td>94.52%</td>
</tr>
</tbody>
</table>

Aggregate by water mains:

<table>
<thead>
<tr>
<th>Water Main No.</th>
<th>Priority</th>
<th>Num SLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WS 1020</td>
<td>Low</td>
<td>0.044</td>
</tr>
<tr>
<td>WS 1244</td>
<td>Medium</td>
<td>1.2</td>
</tr>
<tr>
<td>WS 1234</td>
<td>High</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Phase 2
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

---

**Ability to Predict Material Improves Over Time**

---

Optional: Community-facing map portal
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
Output Data

• Recent SL material confirmation
• Historical SL records
• Parcel attribute data

Input Data

• New SL material verification

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

1

2

3

1

2

3

1

2

3

• Recommended Inspection List
• Region-level preliminary estimates

• Prioritized Replacement List
• System-wide home level inventory

• Continued AI refinement

COPYRIGHTED MATERIAL
VIEW ONLY
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
Can you perform a hydrovac inspection at curb box?

No
- Use other digging method

Yes
- Inspect 1 foot** on each side of curb box
  - Lead or galvanized
    - Needs replacing
      - Lead or galvanized
        - Needs replacing
          - Copper
            - Inspect 2-3 feet more on each side**
              - If in-home inspection revealed fully copper, no further inspection needed.

** Exact numbers determined with city
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.

<table>
<thead>
<tr>
<th></th>
<th>Water test &gt; 0 ppb</th>
<th>Water test = 0 ppb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead Service Line</td>
<td>2935</td>
<td>702</td>
</tr>
<tr>
<td>Non-lead Service Line</td>
<td>2761</td>
<td>2930</td>
</tr>
</tbody>
</table>

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

Flint-Lines Application Overview

Common use cases and instructions
May 15th, 2017 (Updated July 16, 2019)
Initially presented to Flint contractors with Fast Start team in Flint City Hall

Application developed by and presentation prepared by
Alex Chojnacki with Jacob Abernethy, Eric Schwartz, and Jared Webb
Michigan Data Science Team, University of Michigan
Sample size in different-sized communities

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

<table>
<thead>
<tr>
<th>Total number of service lines</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspections in initial set</td>
<td>278</td>
<td>357</td>
<td>370</td>
<td>382</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Total number of service lines</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspections in initial set</td>
<td>400</td>
<td>586</td>
<td>623</td>
<td>655</td>
</tr>
</tbody>
</table>

“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!

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

- **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
- **Cost savings**: $8,300,000
- **Avoided excavations**: 3000
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