Property data in Cleveland/Cuyahoga County, OH

Michael Schramm
Director of Information Technology and Research, Cuyahoga Land Bank
Research Associate, Center on Urban Poverty and Community Development,
Case Western Reserve University
Cuyahoga Land Bank

• Doors opened in June 2009
• Non profit, government purposed community improvement corporation
• Three engines that make Ohio County Land Reutilization Corporations effective:
  – Dedicated funding stream (penalty and interest on collected delinquent property taxes)
  – Enhanced tax foreclosure laws
  – Outside of government (can transact on a dime)
Cuyahoga Land Bank production

- Properties transacted – 6,818
- Properties disposed – 4,567
- Properties in inventory – 2,251
- Properties demolished – 4,656
- Facilitated renovations – 1,397
Center on Urban Poverty and Community Development

Founded in 1988

Our focus:
Neighborhood as a fundamental interface between individuals and systemic forces that drive opportunities.

Specific Goals:
- Through local engagement aim to build knowledge of what works in policy and practice
- Through national partnerships aim to bring community to the forefront of efforts to address social disadvantage (founding partner in National Neighborhood Indicators Partnership housed at the Urban Institute)

A creative team of faculty, staff, and students with diverse set of skills
Community Driven Partnerships
The Center’s areas of work

• Indicator data at the neighborhood level – public portal
• Integrated data at the individual level (persons and properties)
• Research and evaluation projects focused on children, families, and neighborhoods
– Started in 1992, evolved over time
– Data focused at neighborhood level and below
Overview of Center Data

PERSONAL RECORDS
- Birth & death certificates
- Child welfare records
- Juvenile court filings
- Preschool & childcare records
- Home visiting records
- Public school student records
- Public assistance monthly eligibility
- Homeless intake & service records
- County jail admission & discharge
- Lead screenings
- Electronic health records
- Public Housing

PLACE-BASED RECORDS
- Property characteristics
- Sales transactions
- Deed transfers
- Foreclosure filings
- Tax delinquencies
- Postal vacancies
- Lank bank properties
- Parcel shape files
- Crime reports
- Housing code enforcement

CHILD System
- Probabilistic matching & individual record linkage
- Random identifier assigned
- Secure/private

NEO CANDO
- Open tabulated data by block group, census track, & special geographic units (e.g. SPAs, County Districts, Wards, etc.)
- Analysts download data (CSV files) – e.g. researchers, students, agency personal media, consultants
- Display on Google maps

NST Application
- Parcel level data and customized tools for community development sector
- Trained users view customized display
- Users can upload their own data & see it in display
- Filtering, tabulation, reporting tools

External apps
- Customized applications; API enabled.

DUA’s govern use
IRB oversight
Highly secure data transfer & management
There’s gotta be a better way!
NST web app

- Data system of **parcel-level data** for all parcels in Cuyahoga County, OH
- Much of the data is **updated weekly**
- Data is searchable, filterable, sortable, downloadable
- All data are georeferenced into local geographies and target areas
- Quick mapping functions
- User-added data allowed
Houses data relevant for neighborhood stabilization activities by city, county, and community development entities:

- Property transfers
- Tax records
- Foreclosure filings
- Sheriff’s sales
- Cuyahoga Land Bank records
- Cleveland Land Bank records
- Cleveland Building and Housing administrative data
- Cleveland Housing Court administrative data
- Geographic information
- Others
Ways data are acquired

- Email
- FTP
- DVD/CD
- XML Stream
- Dropbox
- Google Drive
- Screen Scraping
- API endpoints

These all occur at various intervals, but most data are updated weekly
NST web app – create data from data

• Vacant lot proxy
  – County bldg value
  – Cleveland demo
  – Private demo permit
  – Cuyahoga Land Bank demo
  – Cuyahoga County Demo Program Demo
Tier 1 – **UNDER CONTROL**
- In Cuyahoga Land Bank Inventory
- Properties pending transfer to Cuyahoga Land Bank
- In Municipal Land Bank Inventory
- State Forfeiture

Tier 2 – **ABOUT TO BE UNDER CONTROL**
- Nuisance Demo (Municipal/Cuyahoga Land Bank)
- Tax foreclosure Affidavit to Municipal Land Bank
- Tax foreclosure Affidavit to Cuyahoga land Bank
- Fannie Mae
- HUD

Tier 3 – **CAN STEER TO PRODUCTIVE USE**
- Tax foreclosure affidavit – not sent to a Land Bank
- Tax foreclosure (not dismissed)
- Tax delinquency

Tier 4 – **OTHER VACANT/BLIGHTED**
- Mortgage foreclosure (not dismissed)
- Bank Owned (not HUD or Fannie Mae)
- Other vacant lot
- Other vacant structure
Boundaries can be dynamically dissolved based on tier and queried. For example, tiers 1, 2, 3 adjacency analysis greater than 2 acres.
Community Development Uses of Data

• Tax foreclosure candidate identification – Cuyahoga Land Bank
• Researching buyers – Cuyahoga Land Bank
• Code Enforcement Partnership – City of Cleveland
• Trend analysis – VAPAC
• Community Control – Housing Court
Tax foreclosure candidate project

• 20,000 tax foreclosure eligible properties in Cuyahoga County that aren’t already in foreclosure
• County Gov’t has resources/capacity to file 4,000 new cases per year
• Which candidates would be most desirable to land banks (vacant lots ➔ City of Cleveland, vacant structures ➔ Cuyahoga Land Bank)
Using NST for tax foreclosure batching

• **First filter** ➔ Fast track eligible and is NOT already in tax foreclosure system (vacant lot/vacant structure)

• **Second filter** ➔ eliminate properties that do not meet land bank acquisition criteria (commercial/industrial/large apartment bldgs)

• **Third filter** ➔ properties land banks want
  – target areas (NSP2, CDC, Economic Development)
  – vacant structures in tipping point neighborhoods/suburbs (potential renovation or resale)
  – land aggregations (tier 3 properties [tax delinquent adjacent to other tax delinquent/tax foreclosure/land bank owned properties “aggregations”])
Property Profile System vs NEO CANDO/NST

• NEO CANDO/NST are planning tools and visualization tools that mash together various public data sources for planning and outreach

• PPS – Land Bank Property Management Database created by the Cuyahoga Land Bank and currently used by 7 land banks to track inventory, attach pictures, create documents, email ticklers, etc.

(Both data systems feed each other dynamically via API endpoints)
Recent study of Cleveland kindergarteners that used CHILD to predict kindergarten readiness based on early childhood exposures and housing.

We built analysis datasets used in studies conducted by Dynometrics:
- Two demolition studies - Demolition raises property values and lowers foreclosure rate.
- Renovation study with Cleveland Neighborhood Progress.

Data Driven Policy Research
We have a bunch of data

• What else should we do with it?????
Beyond Property Surveys

Modeling Vacancy Using Local Administrative Data

April Urban
Center on Urban Poverty and Community Development
The Opportunity

Test the utility of machine-learning techniques on relevant community problem

Make additional use of comprehensive point-in-time survey
What is machine learning?

As explained by a layman

TCI survey is a one-time measure of vacancy. How can we use other available data to help guess a property’s vacancy status?

Using TCI survey as “ground truth”

Comparing to other vacancy estimation methods (postal, REO status)
Have you had any OVV complaints in the last 2 years?

No

Estimate not vacant
94% correct

Yes

What percentage of your taxes have you paid?

<1%

Estimate vacant
79% correct

≥1%

Estimate not vacant
59% correct
Community Use

Single source of information for parcel-level vacancy proxy status (instead of postal vacancy OR vacancy survey OR water shut off)

CDCs that monitor vacancy properties can better target effort

Tested instrument to use when point-in-time survey is outdated
Additional Use

Usable as a measure of vacancy for other research work?

Ideas?
Contact

April Hirsh Urban
Center on Urban Poverty and Community Development
april.urban@case.edu
216-368-3390
Cleveland Property Survey
__________________________
Project Description and Lessons Learned

Frank Ford
Senior Policy Advisor, Western Reserve Land Conservancy
fford@wrlandconservancy.org

September 29, 2016
Two Primary Objectives

Property Vacancy

• How many?
• Where is vacancy concentrated?
• Establish baseline for tracking over time.

*Note: US Postal Data alternative.*

Property Condition

• Rating by condition (A, B, C, D, F)
• Number and concentration by condition?
• Establish baseline for tracking over time.
Scope & Timing

158,000 parcels were surveyed over 16 weeks
• Residential
• Commercial
• Industrial
• Structures as well as vacant land
Survey Team

- 16 Surveyors
- Plus 1 Team Leader
- Surveyors worked in pairs
- Surveyors got one week of training
  - Identifying indicators of vacancy
  - Applying the condition rating scale
  - In Classroom and in the Field
- Project management overseen by Western Reserve Land Conservancy staff
Survey Rating Scale – Sample Conditions

A. Excellent
- No visible signs of deterioration
- Well maintained and cared for
- New construction/renovation
- Historic detailing, unique

B. Good
- Needs basic improvements
- Minor painting
- Removal of weeds
- Cleaning

C. Fair
- Some cracking of brick or wood
- Major painting required
- Deteriorated cornice
- Crumbling concrete
- Cracked windows or stairs

D. Deteriorated
- Major cracking or brick, wood rotting
- Broken or missing windows
- Missing brick and siding
- Open holes

F. Unsafe/Hazard
- House is open and a shell
- Can see through completely
- House ransacked and filled with trash
- In danger of collapse
- Immediate safety hazard to neighborhood
Survey Logistics

• Survey software loaded on IPad Mini Tablets
  – Licensed from ESRI (at beginning)
  – Licensed from Loveland Technologies (later)

• Average survey 3 minutes per house
  – Check off items on IPad
    • Vacancy
    • Specific Itemized Conditions
    • Rating A-B-C-D-F
  – Take photo of house on IPad
  – Data uploaded to server

• Surveyors stay on sidewalk
How The Survey Was Organized

• One neighborhood at a time
• Coordinated in advance with local community organizations
• Teams given maps and assigned to work in sub-areas: census tracts or voting precincts
Two Forms of Quality Control

1. Weekly random samples of each surveyor’s work.
   – Compare photo to rating for consistency

2. All Properties that were found “vacant” and rated “D” or “F”
   – Photo was checked for accuracy and consistency.
Survey Cost

$138,000 Labor
$45,000 GIS Services
$16,000 Equipment & Software
$58,000 Report and Publication
$42,000 Admin

$299,000 Total
Special Considerations

• Consider the season – winter may limit survey opportunities, but ice melt on roof and steam from vents are good indications of occupancy.

• Winter may be safest in high crime areas.

• Summer provides the greatest survey opportunities, but may be less safe; safest times will be early in the AM and mid day; much more street activity in later afternoon.

• E.g., taking a photo of a house with a half dozen people hanging out on the porch. Make a note of those and come back another day.
Limitations of An Exterior Survey With Respect to Vacancy

• When compared to US Postal vacancy data, survey vacancy varied in an interesting way.

• In the most distressed neighborhoods surveyors tended to find slightly higher vacancy than reported by the US Postal data.

• In the most stable neighborhoods surveyors tended to find slightly lower vacancy than reported by the US Postal data.
Limitations of An Exterior Survey With Respect to Rating Conditions

- Surveyors on the sidewalk cannot see wide-open or missing doors in the rear.
- Cannot see conditions on the inside of the house.
- Surveyors found 3,800 vacant Ds and Fs.
- But we learned later that 1,400 other homes rated A, B and C had already been condemned by the City for severe deterioration on the inside.
- In spite of this anomaly, the 3,800 Ds and Fs were invaluable to the City of Cleveland: those that had flown under the radar could now be targeted for inspection.
How the Survey is Being Used
Code Enforcement

Issue or Problem

• Cleveland Building and Housing Department has limited resources to inspect all problem properties.

• Many of the most distressed properties have not yet been identified, inspected or condemned.

Solution

• The Cleveland survey identified the vacant Ds and Fs; City can efficiently prioritize its resources on these.
Minimize Negative Consequences From Bulk Tax Lien Sales

Issue or Problem

• Tax lien buyers bid on a portfolio of liens on thousands of properties; when abandoned properties are discovered, the buyer often chooses to not foreclose and these remain in limbo with a clouded title.

Solution

• The Cleveland survey enables county officials to identify and remove vacant structures from the tax lien sale portfolio; these are redirected to a special Board of Revision tax foreclosure, and conveyed to our county land bank.
Strategic Targeting For Redevelopment

Issue or Problem

• Low median sale prices in our weak real estate market can make the renovation of the most distressed houses financially infeasible.

Solution

• The Cleveland survey enables housing and community development officials to identify the vacant homes that are still in relatively good condition and more viable for redevelopment.
Resource Development and Allocation

Issue or Problem

- Communities with limited resources are often operating “blind”, i.e. where are the vacant homes; where are the vacants that are most distressed; what’s the extent of our problem and how much will it cost to fix it?

Solution

- The Cleveland survey enables housing and community development officials to develop a realistic appraisal of the resources they will need to address the vacant property problem.
Neighborhood Typology

Issue or Problem

• Several years ago Cleveland developed a neighborhood typology that would assist the City in identifying neighborhoods based on trends, e.g. “stable”, “in decline”, “tipping”, etc., but it was based on limited property condition data.

Solution

• The Cleveland survey now enables planners to “ground-truth” the typology with more accurate data.

• The survey provides a baseline for monitoring trends over time.
For Additional Information

Surveys of Ohio cities conducted by the Western Reserve Land Conservancy can be accessed here:
http://www.wrlandconservancy.org/publications-by-type/special-publications/property-inventory-reports/

Frank Ford, Senior Policy Advisor
fford@wrlandconservancy.org
1-216-515-8300
Predictive Modeling of Vacant Properties
Using a Point-in-Time Survey and Periodic Administrative Data

Isaac Oduro, Francisca G.-C. Richter, April Urban
Case Western Reserve University
2016
### Sources of Administrative Data

<table>
<thead>
<tr>
<th>Variable names</th>
<th>variable type</th>
<th>Description</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacant</td>
<td>derived</td>
<td>&quot;Vacant structure&quot; as 1 and 0 otherwise</td>
<td>TCI survey data</td>
</tr>
<tr>
<td>pv_count</td>
<td>derived</td>
<td>counts of postal vacancy since 2006</td>
<td>Postal_vacancy</td>
</tr>
<tr>
<td>al_num</td>
<td>derived</td>
<td>Total number of armsales since 2006</td>
<td>Arms_length_sale</td>
</tr>
<tr>
<td>t_num</td>
<td>derived</td>
<td>Total number of transfers since 2006</td>
<td>Transfer</td>
</tr>
<tr>
<td>al_days</td>
<td>derived</td>
<td>Total number of days since last armsales to present</td>
<td>Arms_length sale</td>
</tr>
<tr>
<td>condition</td>
<td>derived</td>
<td>Categorical: bad, average, good</td>
<td>Building Condition</td>
</tr>
<tr>
<td>days_since_ovv_board_up</td>
<td>derived</td>
<td>number of days since OVV was boarded up since 2006</td>
<td>Complaints</td>
</tr>
<tr>
<td>days_since_ovv</td>
<td>derived</td>
<td>Number of days since parcel has been in ovv since 2006</td>
<td>Complaints</td>
</tr>
<tr>
<td>v_total_1yr</td>
<td>derived</td>
<td>number of violations filed with in 1 yr period before survey date</td>
<td>violations</td>
</tr>
<tr>
<td>v_total_2yr</td>
<td>derived</td>
<td>number of violations filed with in 2 yr period before survey date</td>
<td>violations</td>
</tr>
<tr>
<td>Vacant_block</td>
<td>original</td>
<td>Vacant housing units, percent, 2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>median_rent</td>
<td>original</td>
<td>Median gross rent, number, 2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>property_crimes</td>
<td>original</td>
<td>Property crimes, rate per 100,000 population, 2014 (Crime)</td>
<td>Demography</td>
</tr>
<tr>
<td>burglaries</td>
<td>original</td>
<td>Burglaries, rate per 100,000 population, 2014 (Crime)</td>
<td>Demography</td>
</tr>
<tr>
<td>bachelors+</td>
<td>original</td>
<td>Persons with bachelors degree or more, percent, 2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>poverty_rate</td>
<td>original</td>
<td>Poverty rate, 2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>median_hh_income</td>
<td>original</td>
<td>Median household income, 2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>race</td>
<td>original</td>
<td>2012 5-yr est (ACS 2012 5-year)</td>
<td>Demography</td>
</tr>
<tr>
<td>fc_1yr</td>
<td>derived</td>
<td>Total count of foreclosure filing a year before of survey date</td>
<td>Foreclosure</td>
</tr>
<tr>
<td>fc_2yr</td>
<td>derived</td>
<td>Total count of foreclosure filing 2 years before of survey date</td>
<td>Foreclosure</td>
</tr>
<tr>
<td>active_fc</td>
<td>derived</td>
<td>Sets active foreclosure till date to 1 and 0 otherwise</td>
<td>Foreclosure</td>
</tr>
<tr>
<td>lb_acquired</td>
<td>derived</td>
<td>Encodes acquired parcels as 1 and 0 otherwise</td>
<td>Land bank</td>
</tr>
<tr>
<td>lb_tax_fc</td>
<td>derived</td>
<td>encodes tax foreclosure as 1 and 0 otherwise</td>
<td>Land bank</td>
</tr>
<tr>
<td>total_net_delq_balance</td>
<td>original</td>
<td>Certified total deliquent taxes owed</td>
<td>Tax</td>
</tr>
<tr>
<td>delq_total_ratio</td>
<td>derived</td>
<td>Ratio of tax deliquency to grand total owed</td>
<td>Tax</td>
</tr>
<tr>
<td>paid_percent</td>
<td>derived</td>
<td>Ratio of total paid to grand total owed</td>
<td>Tax</td>
</tr>
</tbody>
</table>
## Logistic Regression Output

Logistic regression  
Number of obs = 88895  
LR chi2(30) = 20722.01  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.3935

Log likelihood = -15970.192

| vacant   | Odds Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------|------------|-----------|-------|-----|----------------------|
| pv_count | 1.203618   | 0.032031  | 42.87 | 0.000 | 1.193463 | 1.213859 |
| property | 0.999959   | 5.756-06  | -0.72 | 0.471 | 0.999859 | 0.999967 |
| tmkval   | 0.999912   | 1.09e-06  | -8.01 | 0.000 | 0.999911 | 0.999913 |
| ownerocd_value | 0.727061 | 0.0247569 | -9.36 | 0.000 | 0.680122 | 0.772395 |
| yrbuilt_filtered | 0.9933569 | 0.0007497 | -8.83 | 0.000 | 0.9891886 | 0.9948273 |
| vacant_block | 1.000824 | 0.0022261 | 0.37 | 0.711 | 0.9967097 | 1.005197 |
| median_rent | 0.999566 | 0.0001595 | -2.79 | 0.005 | 0.9992441 | 0.9998692 |
| property_crimes | 1.000025 | 0.00096e-06 | 2.97 | 0.003 | 1.000009 | 1.000042 |
| burglarita | 1.000049 | 0.0002237 | 2.06 | 0.040 | 1.000002 | 1.000095 |
| bachelors | 1.000612 | 0.0029711 | 2.91 | 0.004 | 1.000296 | 1.014452 |
| poverty_rate | 1.000261 | 0.0023402 | 3.54 | 0.000 | 1.000095 | 1.012805 |
| median_hh_income | 1.000012 | 3.906e-06 | 2.97 | 0.003 | 1.000004 | 1.000019 |
| black | 1.00558 | 0.0005993 | 9.34 | 0.000 | 1.004406 | 1.006755 |
| young | 1.00212 | 0.0031742 | 0.67 | 0.504 | 0.995102 | 1.000361 |
| lb_acquired | 6.480307 | 1.660793 | 7.31 | 0.000 | 3.928726 | 10.71547 |
| lb_tax_fo | 0.3544041 | 0.0860918 | -4.27 | 0.000 | 0.2201531 | 0.570022 |
| total_net_delq_bal-e | 1.000028 | 5.37e-06 | 5.24 | 0.000 | 1.000018 | 1.000039 |
| tax_assessed_land | 1.004765 | 0.0062274 | 0.77 | 0.443 | 0.9926353 | 1.017045 |
| tax_market_land | 0.998313 | 0.0021657 | -0.77 | 0.441 | 0.9940855 | 1.002695 |
| delq_total_Ratio | 3.401977 | 0.2275564 | 18.30 | 0.000 | 2.98397 | 3.87854 |
| v_total_1yr | 1.111337 | 0.0112568 | 10.19 | 0.000 | 1.088869 | 1.134165 |
| v_total_2yr | 1.043 | 0.0064385 | 6.82 | 0.000 | 1.030457 | 1.056966 |
| days_since_cvw_ba-e-p | 0.9948761 | 0.0006999 | -1.24 | 0.215 | 0.993083 | 1.000072 |
| s_cvw_board_num | 6.194244 | 13.94311 | 0.81 | 0.418 | 0.0751539 | 510.5344 |
| days_since_cvw | 0.9993959 | 0.000193 | -31.31 | 0.000 | 0.999381 | 0.999417 |
| c_cvw_num | 3.771828 | 0.744119 | 6.73 | 0.000 | 2.562265 | 5.552388 |
| active_fo | 0.9059538 | 1.113788 | -0.80 | 0.424 | 0.7110412 | 1.154926 |
| fc_1yr | 0.826216 | 0.0892514 | -1.14 | 0.256 | 0.7337569 | 1.025869 |
| fc_2yr | 3.413356 | 0.2107839 | 19.94 | 0.000 | 3.030182 | 3.858501 |
| t_num | 1.222611 | 0.1083664 | 2.28 | 0.023 | 1.0286298 | 1.455533 |
| _cons | 84.9251 | 126388.3 | 7.63 | 0.000 | 4603.208 | 1066808 |
Comparing Performance of Models

- False Positive: Predict vacant when not.
- True Positive: Correctly predict vacant.
- Want highest rate of True Positives for a given –low level- of False Positives.
Comparing Performance with Imperfect Data

<table>
<thead>
<tr>
<th>Survey Vacancy</th>
<th>Model Vacancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a b</td>
</tr>
<tr>
<td>0</td>
<td>c d</td>
</tr>
</tbody>
</table>

Survey: occupied
Model: vacant
"FALSE POSITIVE"

Survey: vacant
Model: occupied
"FALSE NEGATIVE"
Patterns of “False Positives”

**Random Forest - Rate of False Positives (%)**

- 0.00
- 0.01 - 1.60
- 1.61 - 2.58
- 2.59 - 9.99
- Census Tracts - no data
- Cuyahoga County Subdivisions
- City of Cleveland

**Gradient Boosting - Rate of False Positives (%)**

- 0.00
- 0.01 - 1.38
- 1.39 - 2.84
- 2.85 - 11.11
- Census Tracts - no data
- Cuyahoga County Subdivisions
- City of Cleveland
Pattern of “False Negatives”

Random Forest - Rate of False Negatives (%)
- 0.00 - 2.13
- 2.14 - 4.94
- 4.95 - 7.69
- 7.70 - 17.70*
- Census Tracts - no data
- Cuyahoga County Subdivisions
- City of Cleveland

*Four highest false negative rates were eliminated as outliers

Gradient Boosting - Rate of False Negatives (%)
- 0.00 - 2.08
- 2.09 - 4.35
- 4.36 - 6.92
- 6.93 - 15.08*
- Census Tracts - no data
- Cuyahoga County Subdivisions
- City of Cleveland

*Four highest false negative rates were eliminated as outliers