Integrating Open Source Data into Utility Customer Systems

Rich Crowley, DNV GL, Portland, ME (*presenting*)

Riley Hastings, Eversource, Westwood, MA

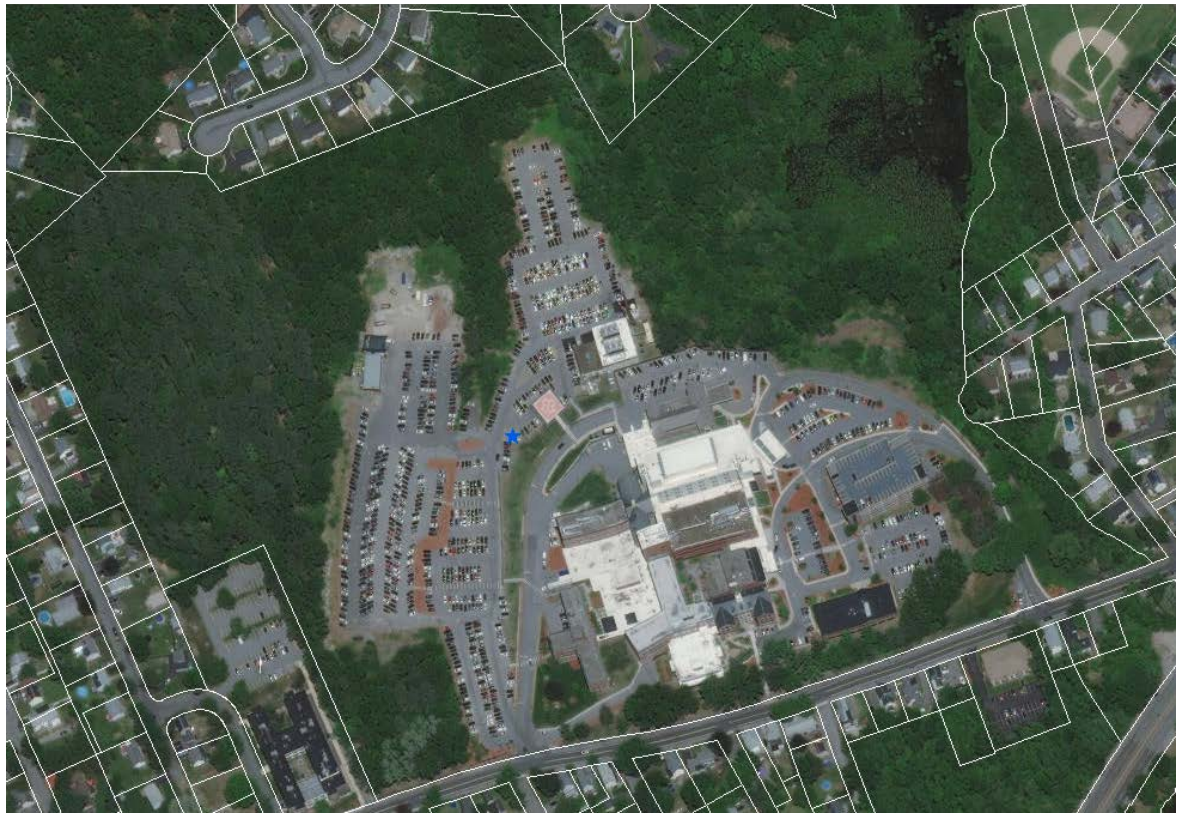
Steve Menges, National Grid, Waltham MA

Bob Wirtshafter, Wirtshafter Associates, Inc., Chatham MA

2019 International Energy Program Evaluation Conference, Denver, CO

Discussion Overview

01	Introduction and background
02	Data integration: sources and methods
03	Data integration: challenges and solutions
04	Locational analytic targeting
05	Customer analytic targeting



How do we get data talking when there is no clear link between datasets?

Background: Why we did this, and why it matters to you

"***trying to make connections in our data jungle is next to impossible sometimes***"...why?!

- Systems do not talk / share any similarities
- Institutional knowledge is inaccessible (or retiring!)
- Too much material, and too complex
- Help is not affordable
- Rationalization: data would not tell us anything new

Massachusetts – 5+ years data, 7 program administrators, legacy company IT systems.

the "data jungle" does not have to be daunting; it also does not have to be complex, expensive, or out of reach

Data Used

Thematic Data

- Location - primary shared grain
- Text – non-standard, but (usually) very descriptive
- Descriptive attributes – refine matches and confidence

Data you (probably) have

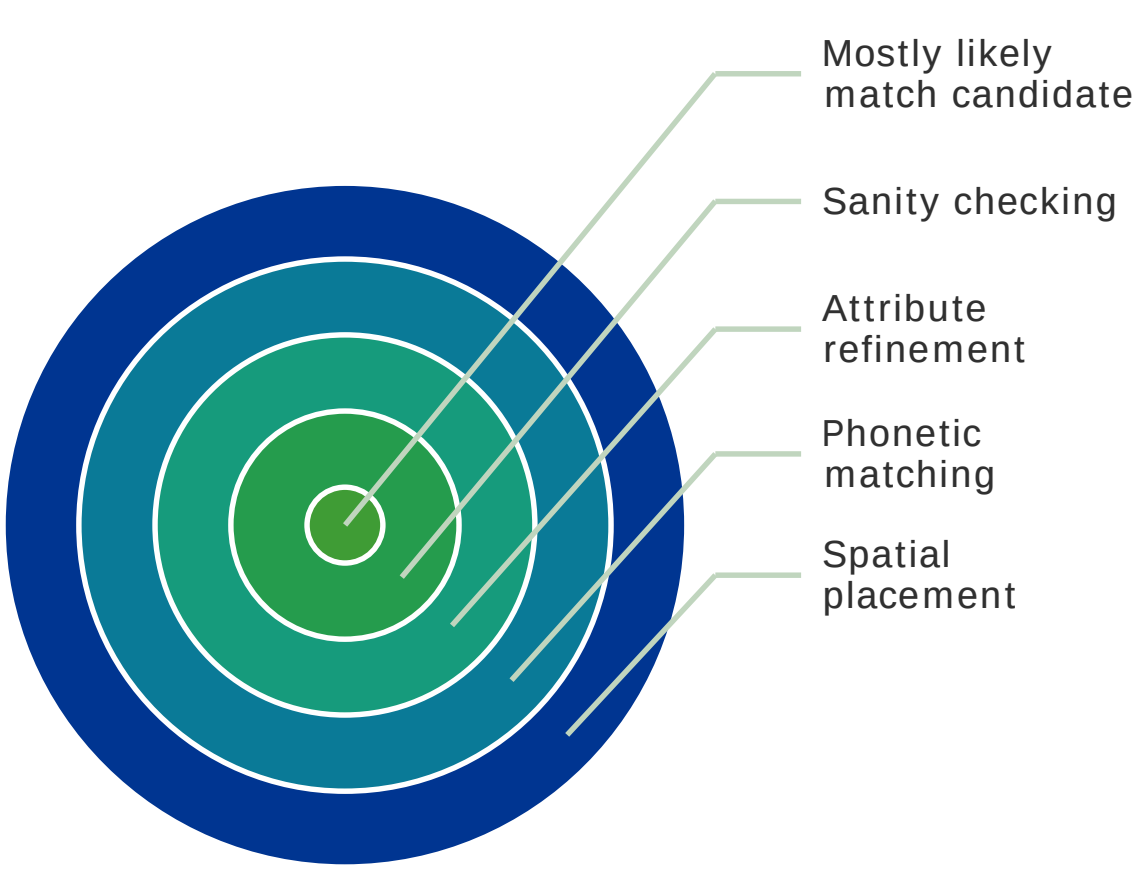
Specific Data Sources (*paper table 1*)

- PA Billing and Tracking inventory data
- Tax assessment building data (MA L3 Parcels)
- EPA emissions and industry data (FRS database)
- Geographic summary data (e911, ACS, ZCBP, URCC, LULC)

The data to integrate

You probably already have a lot of this information… or at least a quick link to it!

Methods, Challenges, and Solutions



Spatial Placement

- Address records are inaccurate
- Address is missing
- Geocoder is inaccurate
- Data grain is inconsistent

Phonetic Matching

- Names inconsistent
- Names missing
- Names unstructured

Attribute Refinement

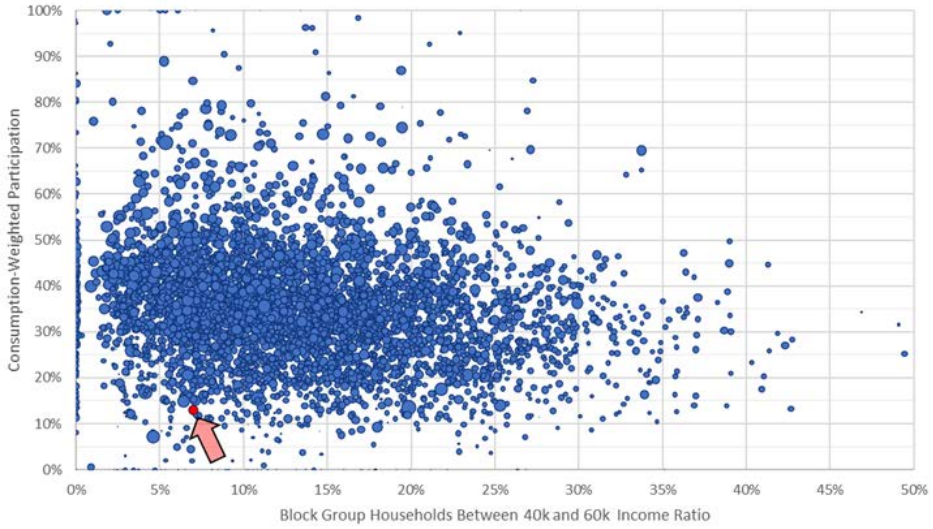
- Phone number alignment
- Consumption relative to building size
- Neighbourhood checks

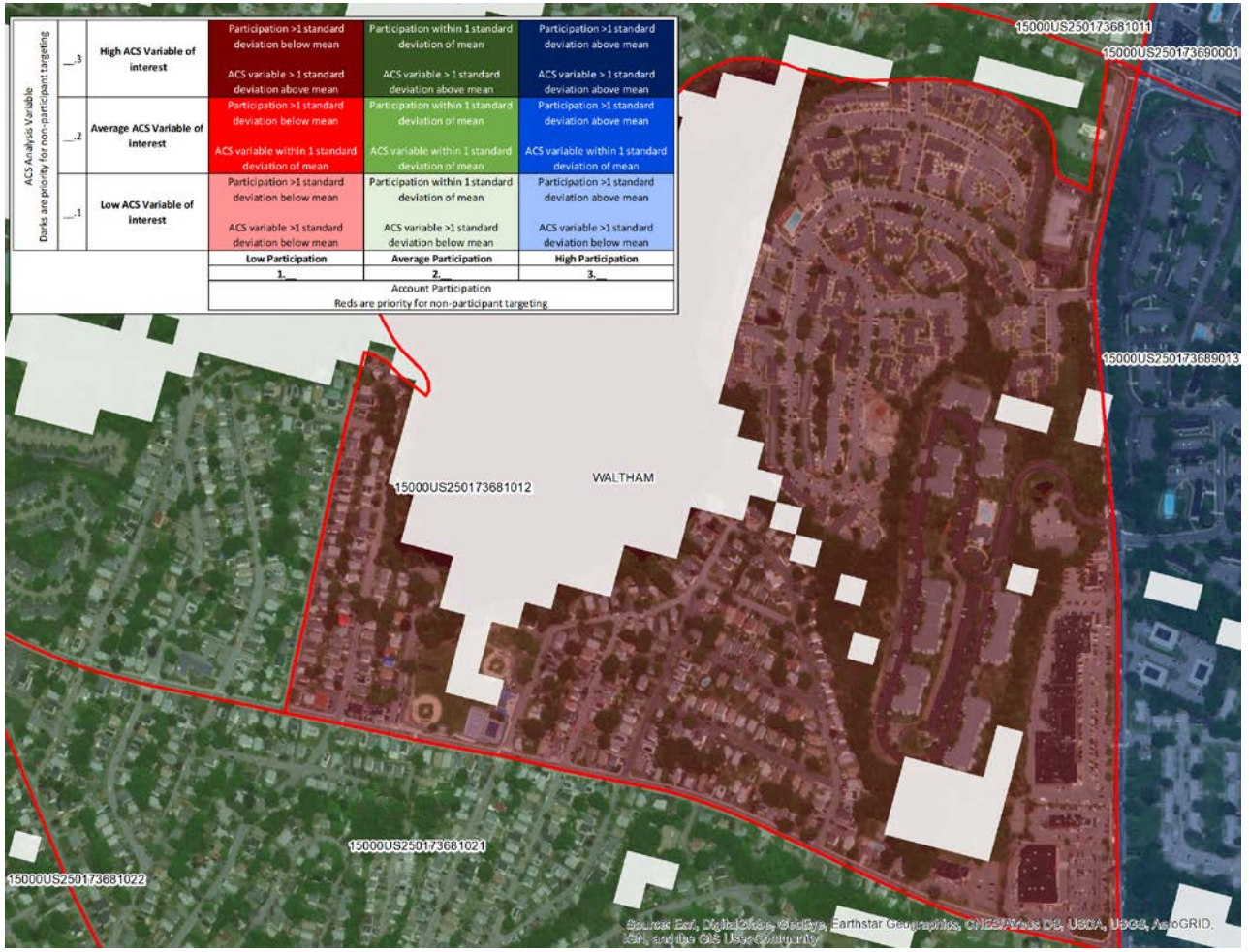
For best results – especially point level – there is no one silver bullet: use everything you have!

Locational analytic target

- Bivariate map – moderate income to unique street address participation rate
  - Z-scores to identify outliers
  - Mask out land use

- Identification of comparable groups



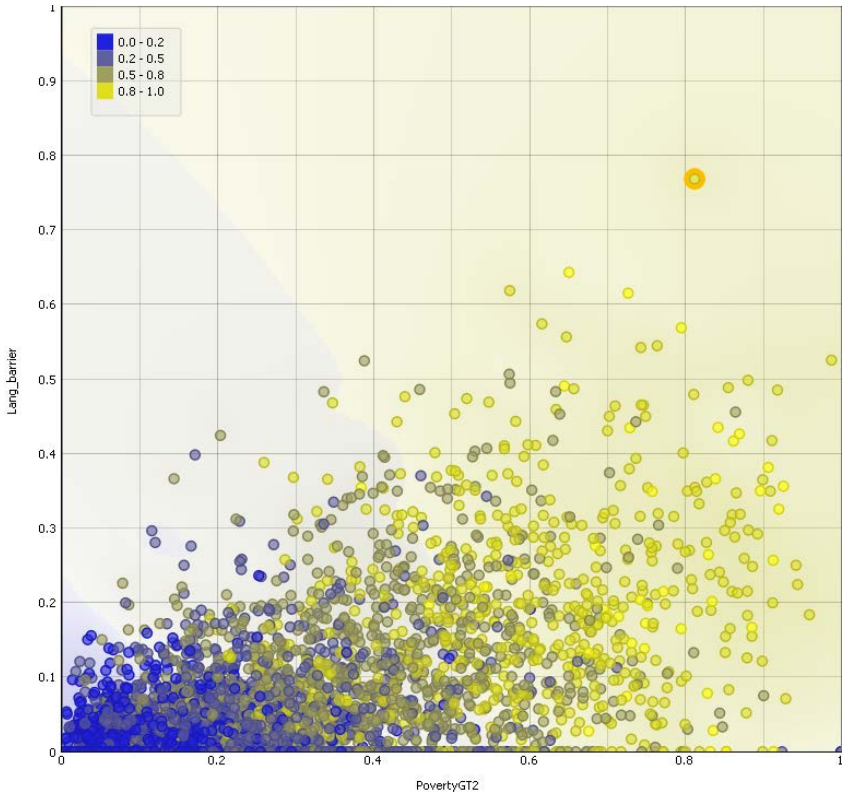


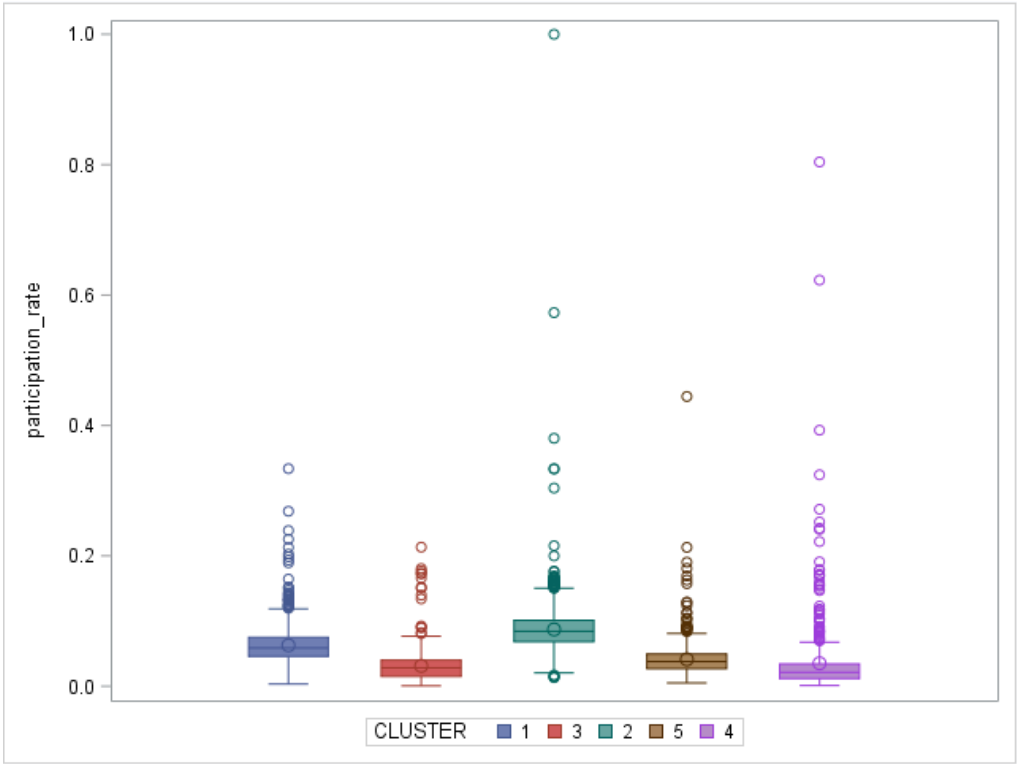
Customer Analytic Targets

- Town homes… same vintage, style, and sizes… but not same participation and savings
- Single family cross fuel opportunities
- Affordable housing units in the apartment complex
- Using different program tracks… can impact savings and measure choices
- Tax data can help understand ownership structures! (Owner-occupied, LLC, renters, etc.)

What the future holds

- Geography: critical for data integration, and for data insights
- Once integrated… how to leverage?!
- Real time or informative time… real time is not always feasible or desirable!





For any questions please contact:

Richard.Crowley@DNVGL.com

+1 207 274 1231

www.dnvgl.com

SAFER, SMARTER, GREENER