Using GIS to improve Addressing

International Geospatial Geocoding Conference

Mark Greninger
Geographic Information Officer
County of Los Angeles

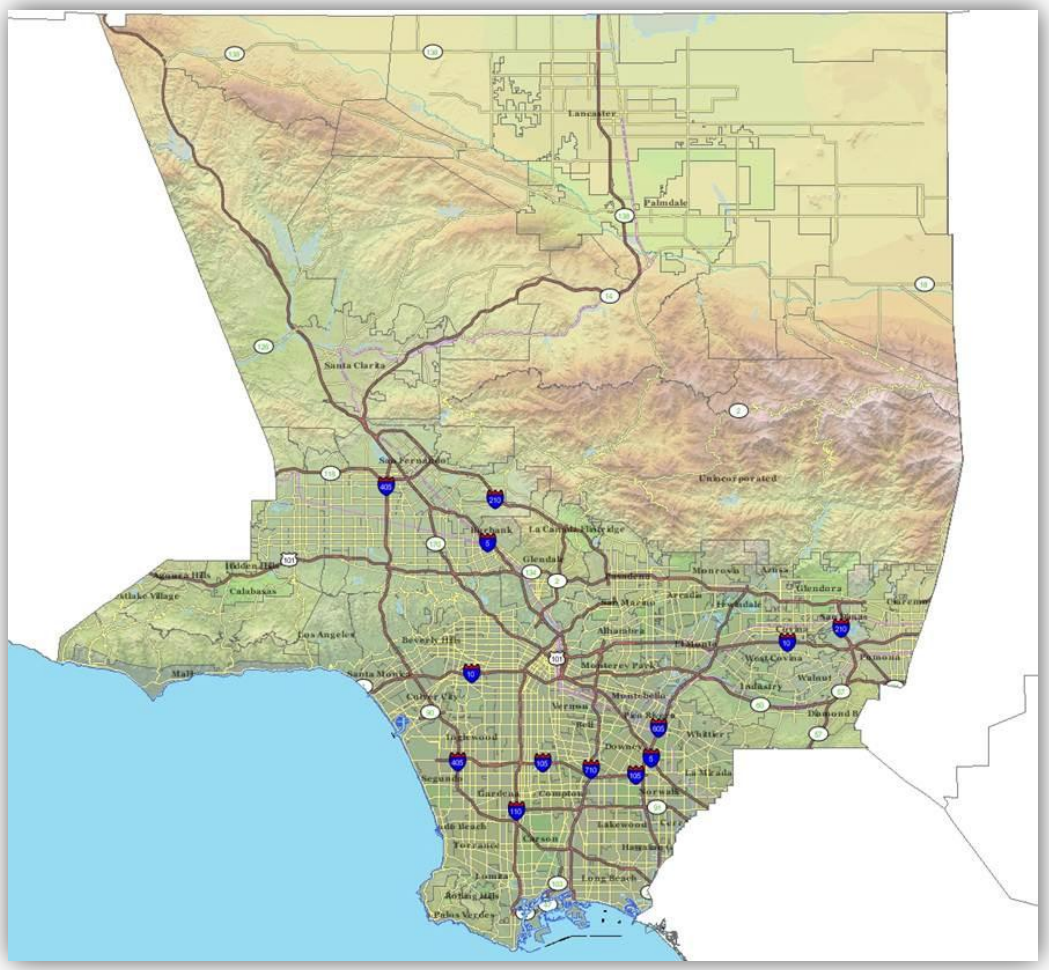
December 7, 2011

Agenda

- A Brief History of LA County Geocoding
- LA County Address Systems
- Using GIS to improve addresses
- Addresses in the future

Los Angeles County

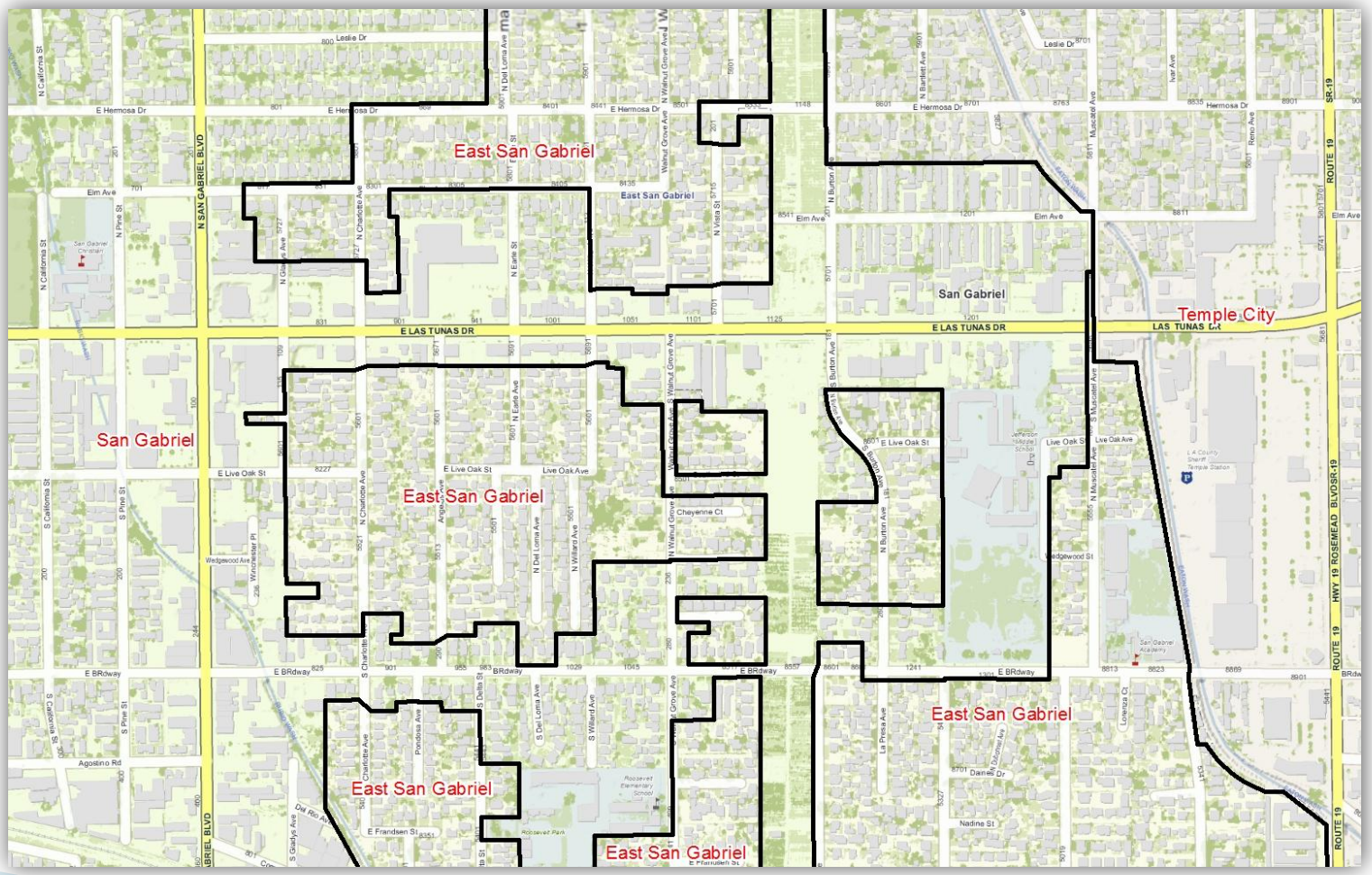
A Map


Lancaster
Palmdale
Santa Clarita
San Fernando
Unincorporated
Burbank
La Canada Flintridge
Glendale
Pasadena
Monrovia
Azusa
Glendora
Hidden Hills
Agoura Hills
Calabasas
Malibu Village
Los Angeles
Beverly Hills
Arcadia
San Marino
Alhambra
Covina
San Dimas
Pomona
West Covina
Walnut
Industry
Diamond B
Monterey Park
Montebello
Santa Monica
Culver City
Vernon
Bell
Pico Rivera
Whittier
Inglewood
Downey
La Mirada
Norwalk
Segundo
Gardena
Compton
Lakewood
Cerritos
Hawaiian
Carson
Torrance
Lomita
Rolling Hills
Palos Verdes
Long Beach

LA County Facts

- Geography
  - 4,084 square miles
  - 88 cities
  - 9.8 million residents
    - 1 of every 4 Californians
    - 1 of every 30 Americans
    - 8th largest state in the US
    - 4 million voters
    - 1.2 million social service recipients
  - More than 60 different jurisdiction types
    - School Districts
    - Water Districts
    - Police/Sheriff
  - More than 30 addressing systems

My favorite city Boundary



How we use addresses

- Validating vs. Geocoding
  - Validating does not allow incorrect addresses
  - Geocoding tries to find the best match
- Dynamic vs. Automated
  - Returning multiple addresses
  - Returning the single best/authoritative address
- Very different business rules

Where we use addresses

- Elections (managing 4 million voters)
- Citizen response (identifying jurisdictions)
- Mapping crime (> 250,000 reports/year)
- Welfare (1.2 million recipients)
- Permitting
- Dispatching (> 500,000/year)
- Enhancing revenue (Census/LUCA found 80,000 new addresses = $160 million over 10 years)
- Emergency Response (Station Fire/Reverse 911)
- Validating Addresses (> 200,000 Child Abuse Allegations/year)

A Brief History

- Geocoding in LA County

My 2nd favorite boundary

LA County Address Data

- 1970s – Supported the development of the TIGER file
- 1980s – Assessor uses GIS to maintain parcels
- 1990s – License Thomas Brothers
- 2000s – begin updating Thomas Brothers with TUS program.
- 2006 – Countywide Address Management
  - Brings points and street lines together
- 2012 – Moving to TIGER

CAMS

- Countywide Address Management System
  - Collaborative Approach to managing Addresses
  - Data owners update their own information
  - 88 cities, 30 departments, and more …
- What is it?
  - Database to store addresses
  - Software to manage them
  - Application to provide feedback
  - Reference File product (*the CAMS data*)
  - Program to govern it

LA County Address Matching

- 1980s – AutoStan/AutoMatch
  - Developed by Matt Jaro
  - Licensed by ESRI and part of ESRI product suite.
- 1990s
  - LA County geocodes with Automatch
  - Automatch acquired by various companies – ends up at IBM in late 1990s
- 2009
  - Merged the forks back together (LA County's matching rules inside ESRI)
- 2011
  - ESRI moves to new matcher in ArcGIS 10

An Address System

Address System Components

- Three related components
  - Reference File
  - Geocoders (Standardizing & Matching)
  - QAQC

Reference File

- Authoritative Address list
  - Street Ranges
  - Address Points
- LA County Sources
  - Thomas Brothers (moving to TIGER)
  - Post office USPS zip+4 file
  - Parcels (maps and database)
  - City Address files
  - Points of interest

Geocoders

- Standardizing rules
  - Breaks addresses into component parts (**correctly**)
  - Recognizes abbreviations and colloquialisms
  - Recognizes patterns and frequencies

Specialized Standardization

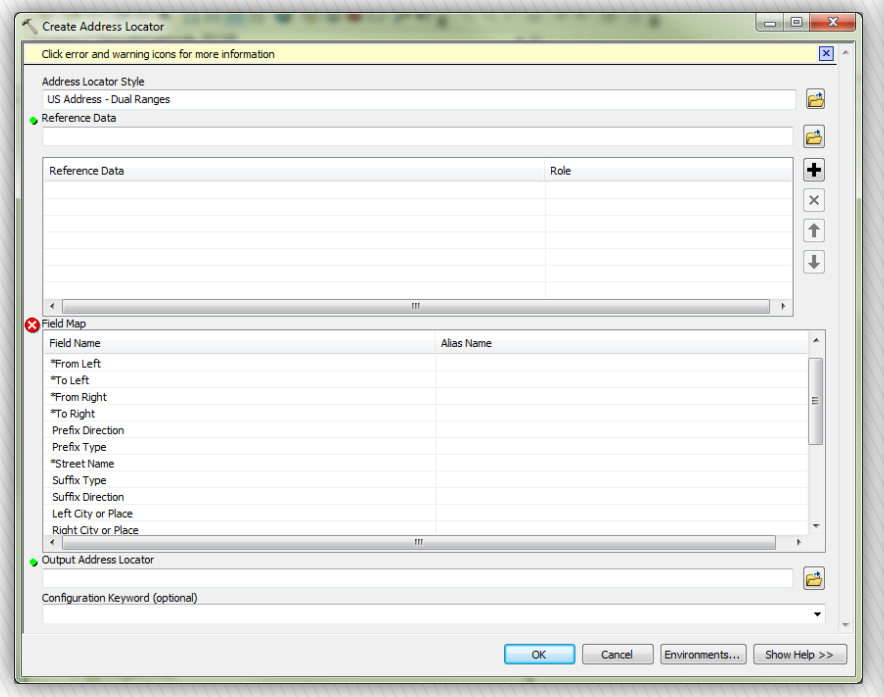
- House Number
- Prefix Direction
- Prefix Type
- Street Name
- Street Type
- Suffix Direction
- Zone

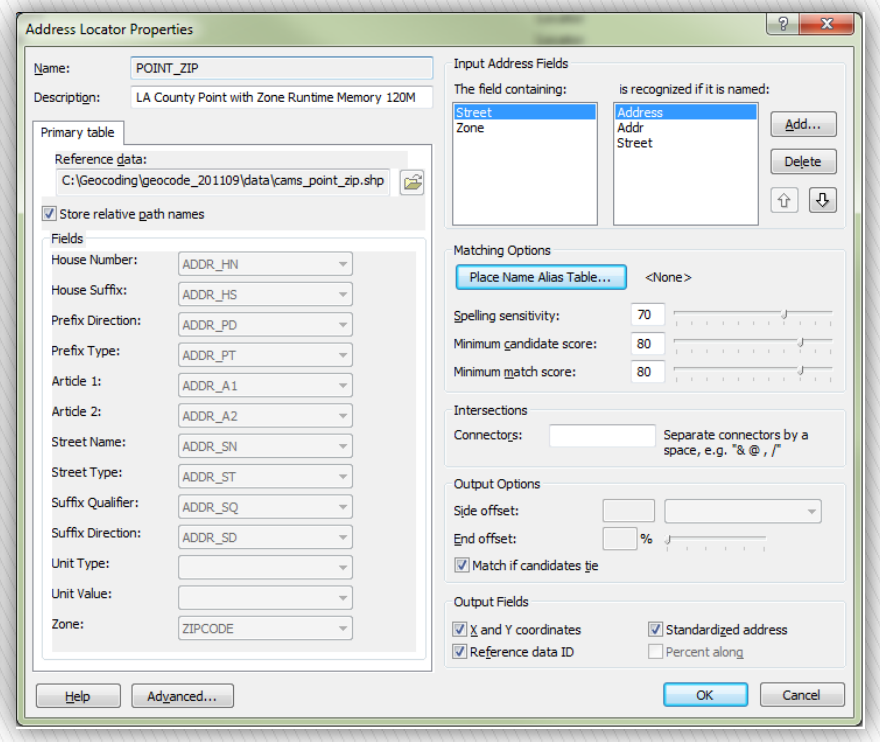
US Street Address Components

- House Number
- **House Suffix (*1/2, A)***
- Prefix Direction
- Prefix Type
- **Article 1 *(of, de)***
- **Article 2 (*the, la)***
- Street Name
- Street Type
- Suffix Direction
- **Unit Type (*Suite, Apt)***
- **Unit Value *(#101, B)***
- Zone

LA County Address Components

Specialized Standardization





US Address – Dual Ranges

LA County Points

Some Examples

- Streets with “The”
  - THE OLD ROAD, THE VILLAGE, THE PROMENADE
- Business roads
  - CENTER POINT PARKWAY
  - OUTER TRAFFIC CIRCLE
- Street with two articles
  - AVENUE OF THE STARS
- Spanish
  - LOS COYOTES DIAGONAL
  - PASEO LOS MONTEROS, VIA LOS MIRADORES, AVENIDA LOS PALOS, LOS CALLE ELEGANTES
  - VISTA DEL VALLE DRIVE
  - PLAZA DEL AMO

LA County's Rules

- Will be available after the conference
  - Only works with ArcGIS 9.3 (not 10)
- LA County GIS Data portal:

http://gis.lacounty.gov/dataportal

Matcher

- Matching rules
  - Assigns weights to components
  - Leverages fuzzy logic to match incorrect names
    - Soundex
    - NYSIIS
    - Reverse Soundex, Reverse NYSIIS, etc
- We have developed special weighting.
  - Lower weights for intersection matches

QAQC

- Some ideas
  - Get rejects rates from an authoritative test file
    - Should be geographically comprehensive
    - Many types of addresses (residential, commercial)
    - If possible get addresses with XY coordinates (parcels)
    - Should have real addresses with errors
  - Geocode one reference file against another
  - Standardizer test (compare the before and after)
- The most difficult test – false matches

False Matches

- Rejects
  - The “known unknown”
  - “Unknown known” when you have matcher failures
- False Matches – the “unknown unknown”
  - Make your matching seem better than it is.
  - Incorrectly assign resources and endanger lives.
  - LA County had a 2.8% false match rate.
- How can we fix false matches?

Using GIS to Improve Addresses

- The Map is the key

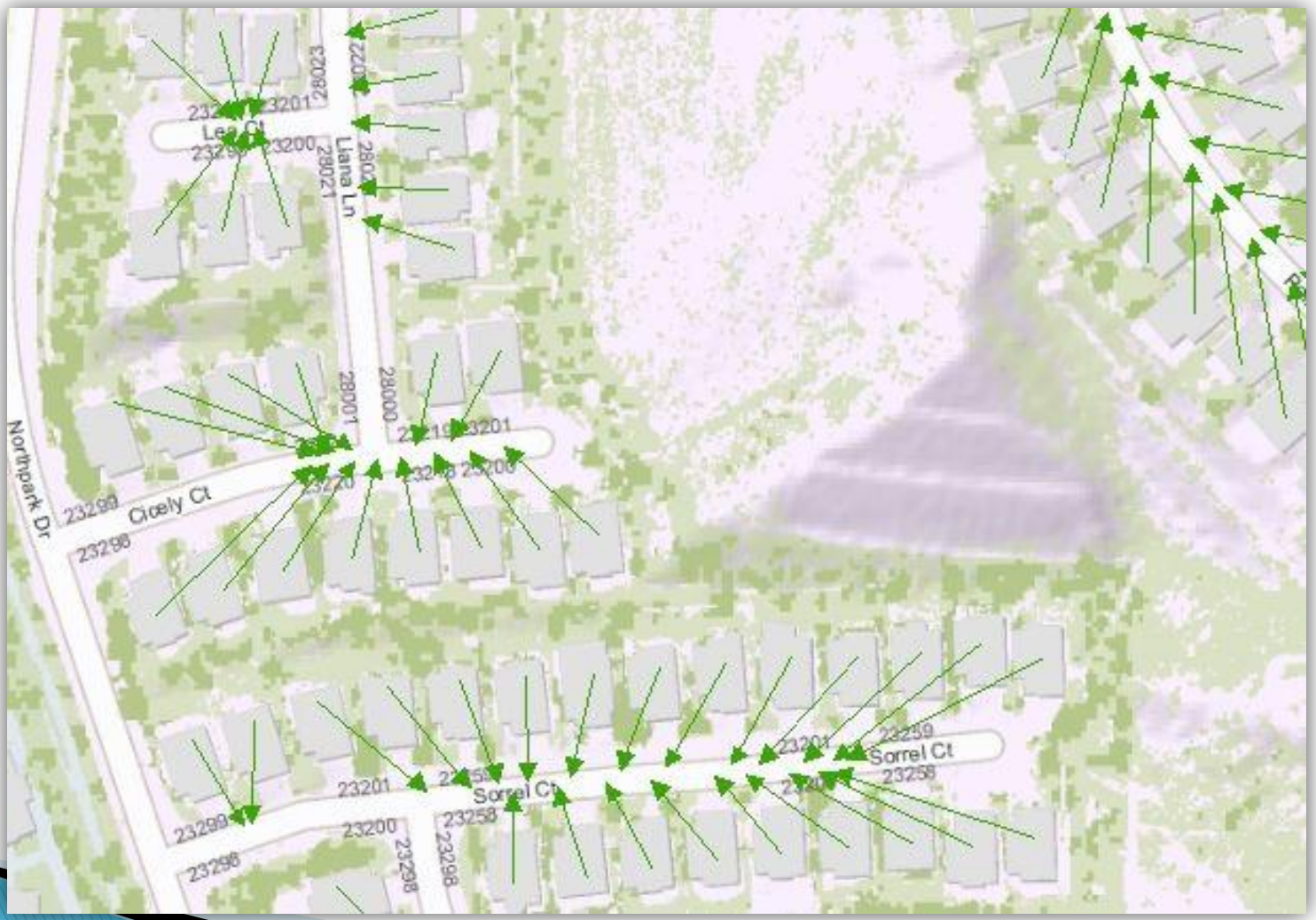
Where are your addresses?

http://gis.lacounty.gov/gisviewer

QAQC – Finding False Matches

- Compare where geocodes should be with where they are.
- How? **Fishbones**
  - Take an authoritative address source and geocode it against another file.
  - Draw a line from the source location to the geocoded location

Fishbones


Lea Ct
Liana Ln
Northpark Dr
Cicely Ct
Sorrel Ct
Sorrel Ct
23201
23200
23298
23299
23258
23259

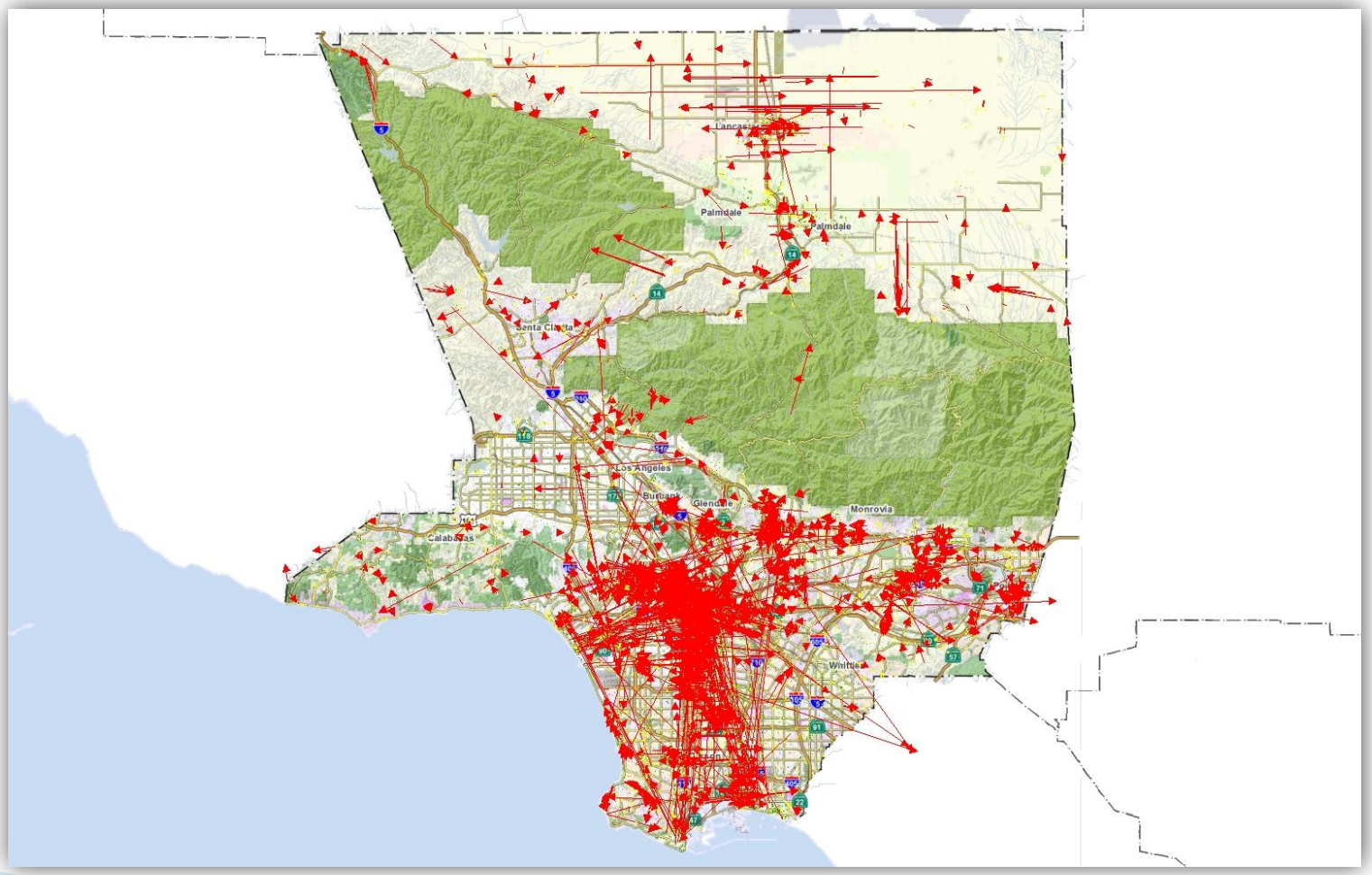
Fishbone Rules

- Rules
  - Lines should not cross streets
  - Lines should not cross each other
  - Lines should not be too long
- GIS automates error finding
  - GIS tool: **intersect/select by location**
  - Intersecting lines have errors.
    - Intersecting with each other
    - Intersecting with the other streets.

Cartographic Aids

- Visual patterns
  - Explosions
  - Implosion
  - Hops
  - Stretches
  - Spiders/Teepees
  - Magnets
  - Crossers
  - Rockets
- Color by line length
  - Short are green < 500 feet)
  - Medium are yellow (500 – 1,000 feet)
  - Long are red (> 1,000 feet)

LA County Fishbone (61,000 errors)



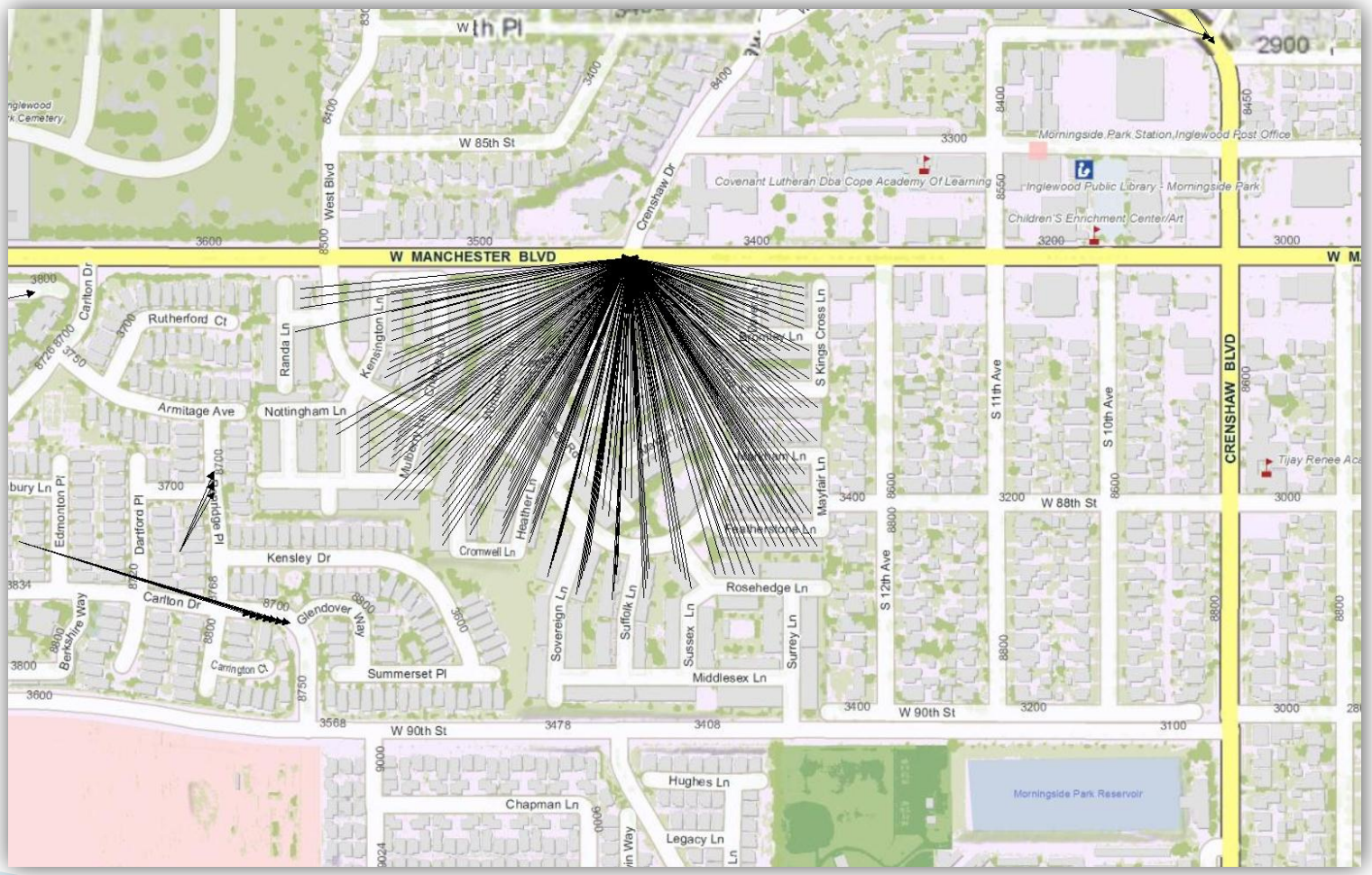
Explosions

Santa Clarita
Esperanza Way
Landmark Way
Northpark Dr
Daydream Way
Reunion Ct
Cloely Ct
Sorrel Ct
Aspen Ridge Dr
Northpark Elementary School
Cheseborough Park

Explosion

- Source information has one xy coordinate for many addresses.
- Not an address reference file error
- Points to opportunity to improve source file (assessor).

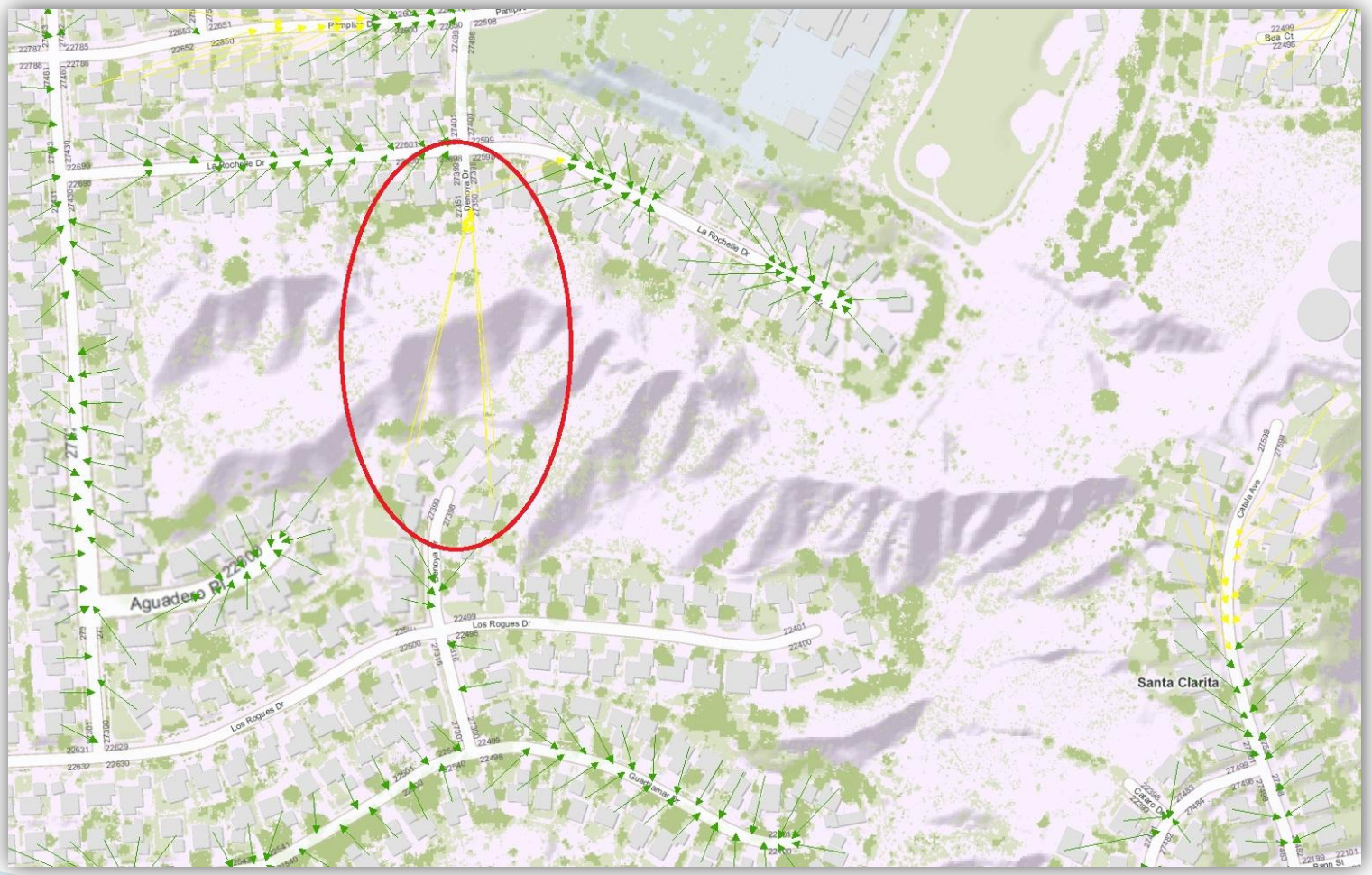
Implosions


W MANCHESTER BLVD
W 85th St
W 88th St
W 90th St
CRENSHAW BLVD
West Blvd
Crenshaw Dr
Rutherford Ct
Randa Ln
Kensington Ln
Armitage Ave
Nottingham Ln
Carlton Dr
Edmonton Pl
Dartford Pl
Kensley Dr
Glendover Way
Berkshire Way
Carrington Ct
Summerset Pl
Cromwell Ln
Heather Ln
Sovereign Ln
Suffolk Ln
Sussex Ln
Surrey Ln
Middlesex Ln
Rosehedge Ln
Mayfair Ln
S Kings Cross Ln
S 11th Ave
S 10th Ave
S 12th Ave
Chapman Ln
Hughes Ln
Legacy Ln
Morningside Park Reservoir
Morningside Park Station,Inglewood Post Office
Inglewood Public Library - Morningside Park
Children'S Enrichment Center/Art
Covenant Lutheran Dba Cope Academy Of Learning

Implosions

- Source information has many points with a single address.
- Condos, large units, etc.
- **Not** necessarily a reference file error
- Could be an error with sub-address information (units, suites, etc.)

Hops



La Rochelle Dr
Los Rogues Dr
Aguadero Pl
Santa Clarita
Catala Ave
Bea Ct

Hops

- Hopping over areas that have no addresses
- Shows street segments with incorrect range assignments

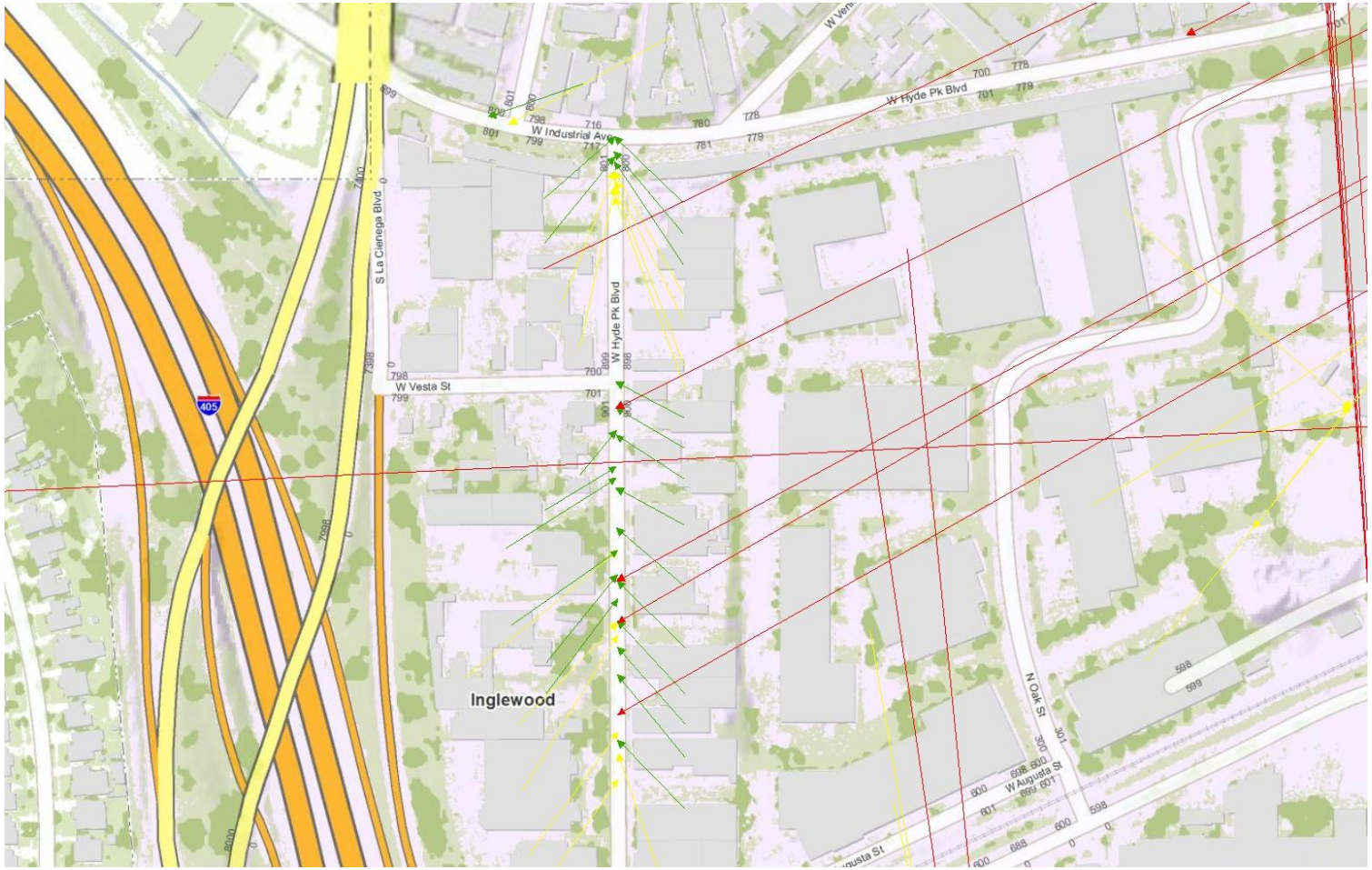
Stretches

BEVERLY BLVD
West Hollywood
Los Angeles
Miracle Mile
W 3RD ST
S ROBERTSON BLVD
N ROBERTSON BLVD
Alden Dr
N Almont Dr
N La Peer Dr
N Swall Dr
N Clark Dr
N Wetherly Dr
S Almont Dr
S Swall Dr
S Clark Dr
S Wetherly Dr
Alley
Bonner Dr
BEVERLY

Stretches

- Street Segments with incorrect range assignments
- Check for duplication or range shifting

Switches


W Hyde Pk Blvd
W Industrial Ave
S La Cienega Blvd
W Hyde Pk Blvd
W Vesta St
405
Inglewood
N Oak St
W Augusta St

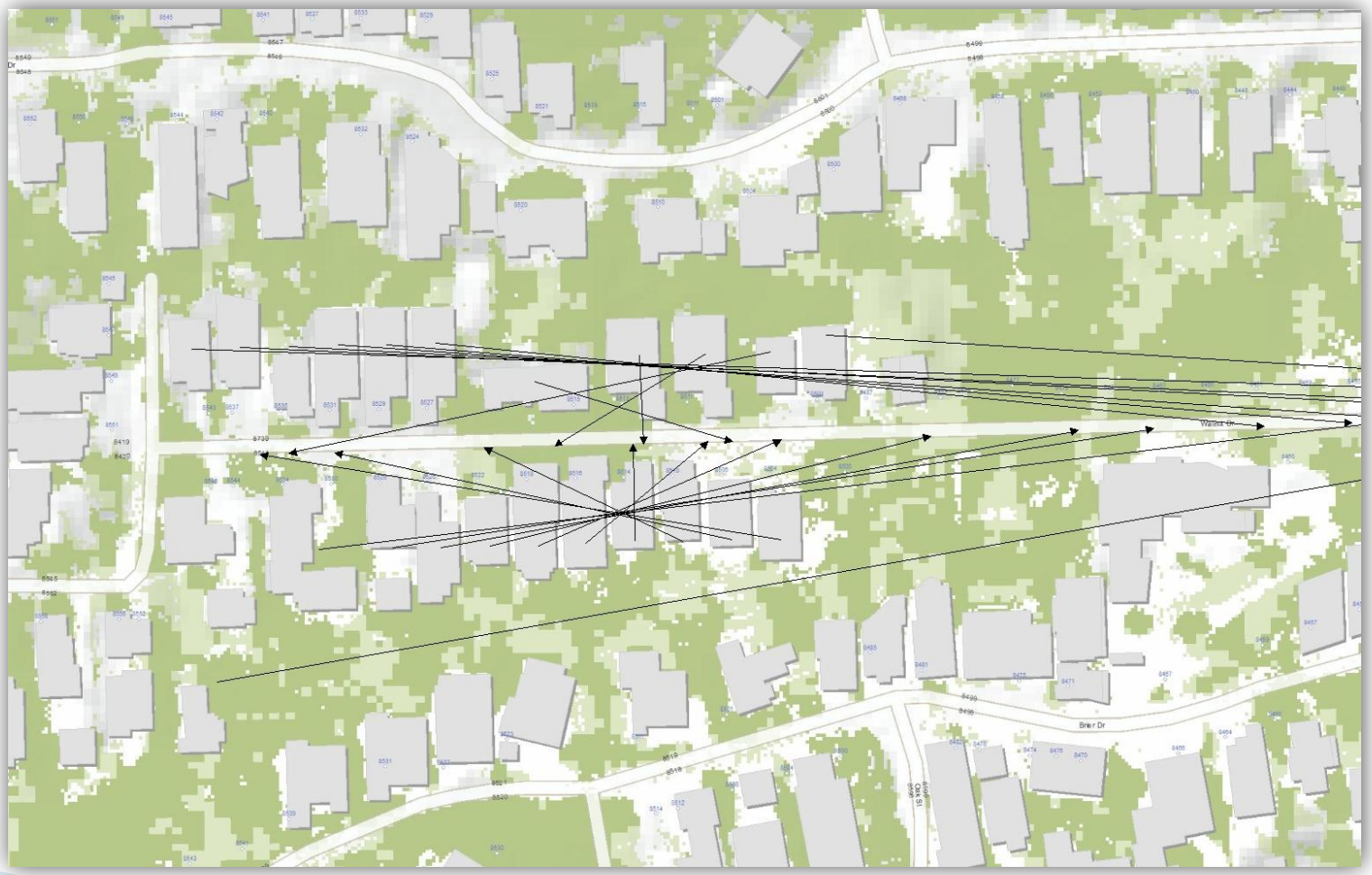
Switches

- Odd and Even sides are switched
  - Left should be odd but it's even
  - Right should be even but it's odd.
- Don't intersect each other but intersect streets.

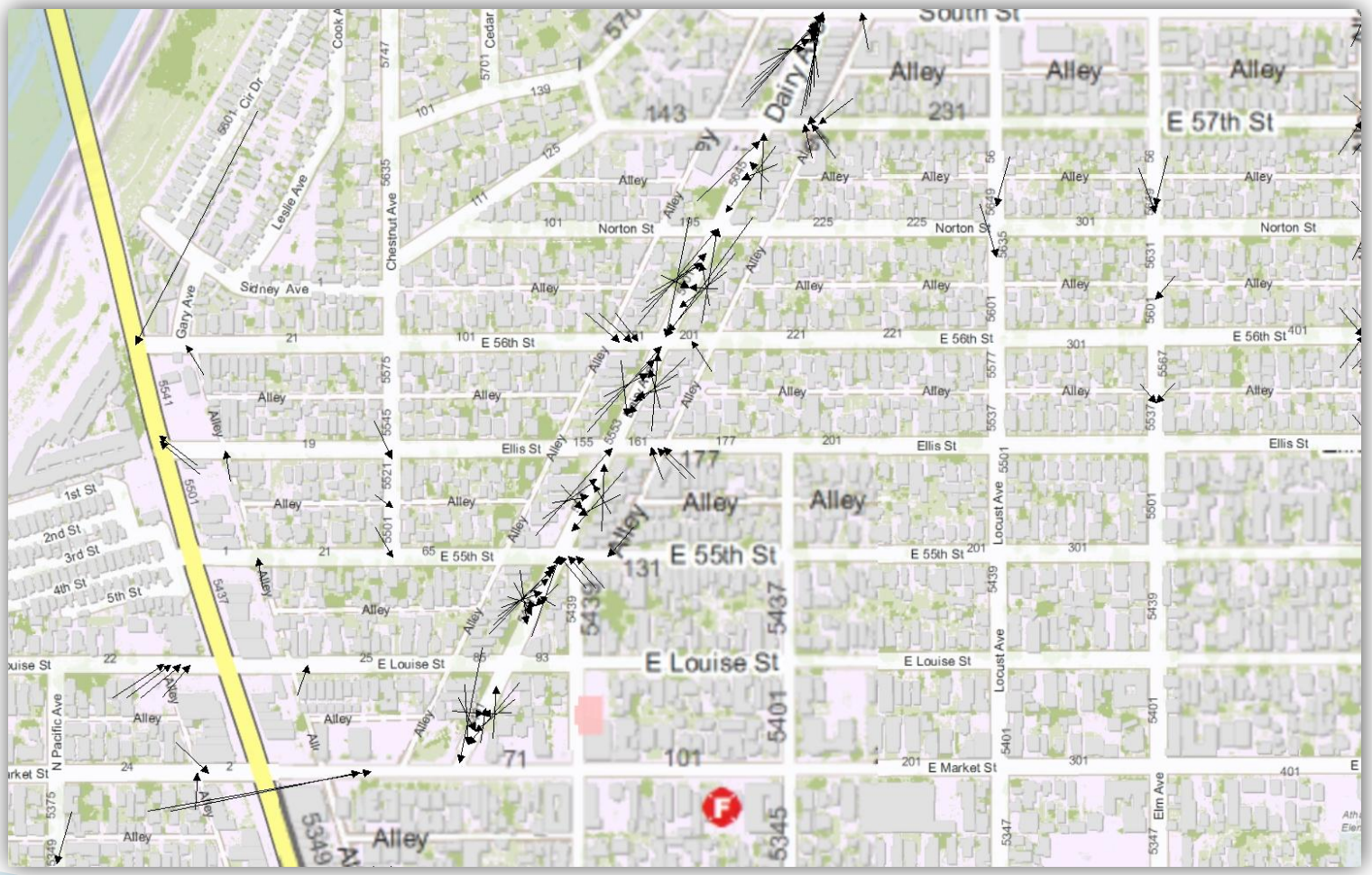
Spiders

2798
2700
2698
2600
2500
2499
2498
2199
2198
2196
N Salinas Ave
N Dorn Ave
Riddle Ave
N Corlett Ave
Compton
Mckinley Ave

Teepees



Spiders/Teepees



Spiders/Teepees

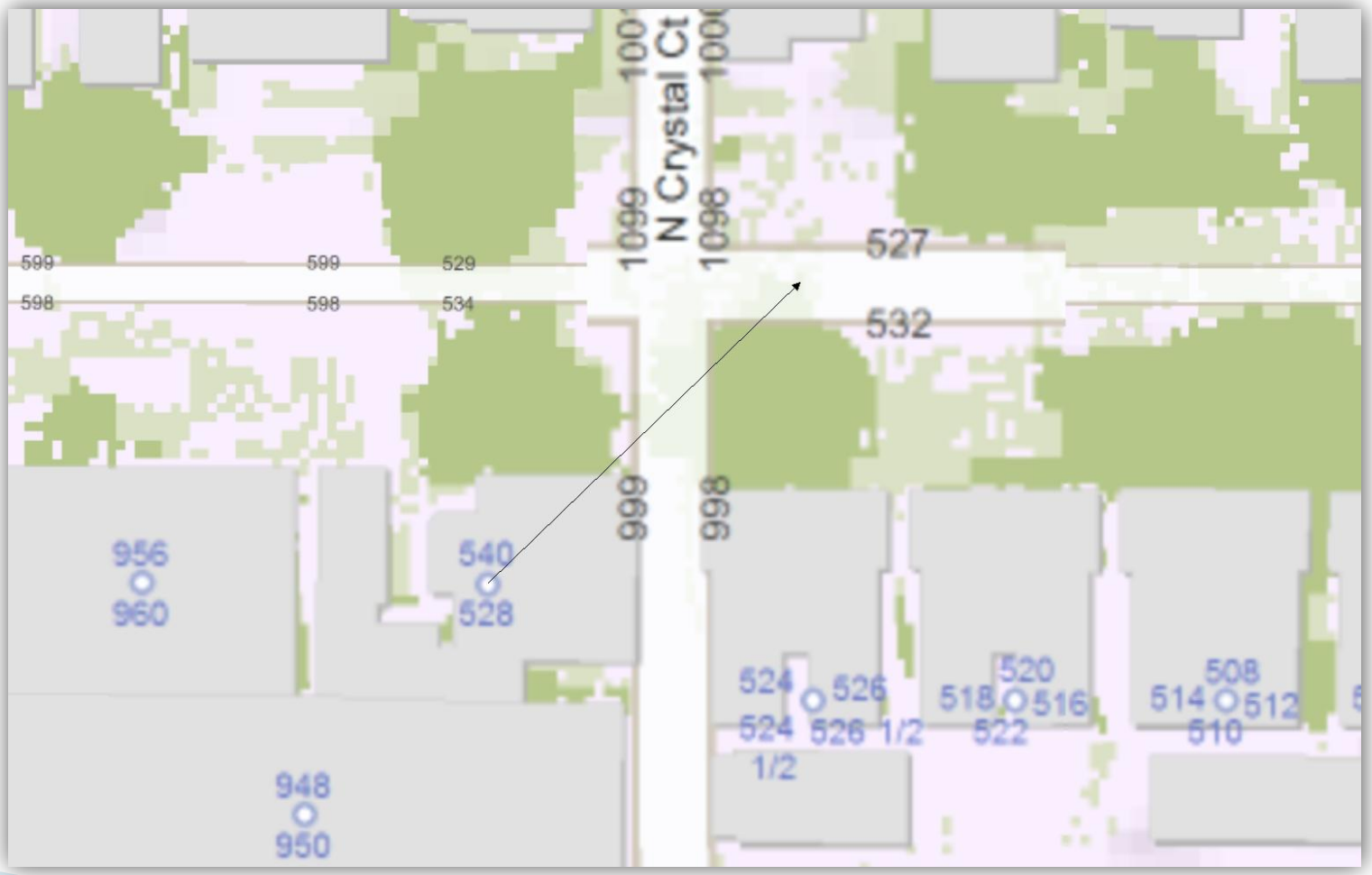
- Flipped range assignments caused by one of two problems
  - Low and high are flipped
  - Street direction is flipped
- These errors intersect with each other

Magnets

Magnets

- Addresses from one street attracted to another
- Can Show a matcher error
- “La Brea” is matching to “La Palma”
- Can’t handle Spanish names (Buena Vista, etc)
- This is a more complicated fix – need to investigate the matcher
- Could be the standardizer
- Could be the matching algorithm

Corner Crossers

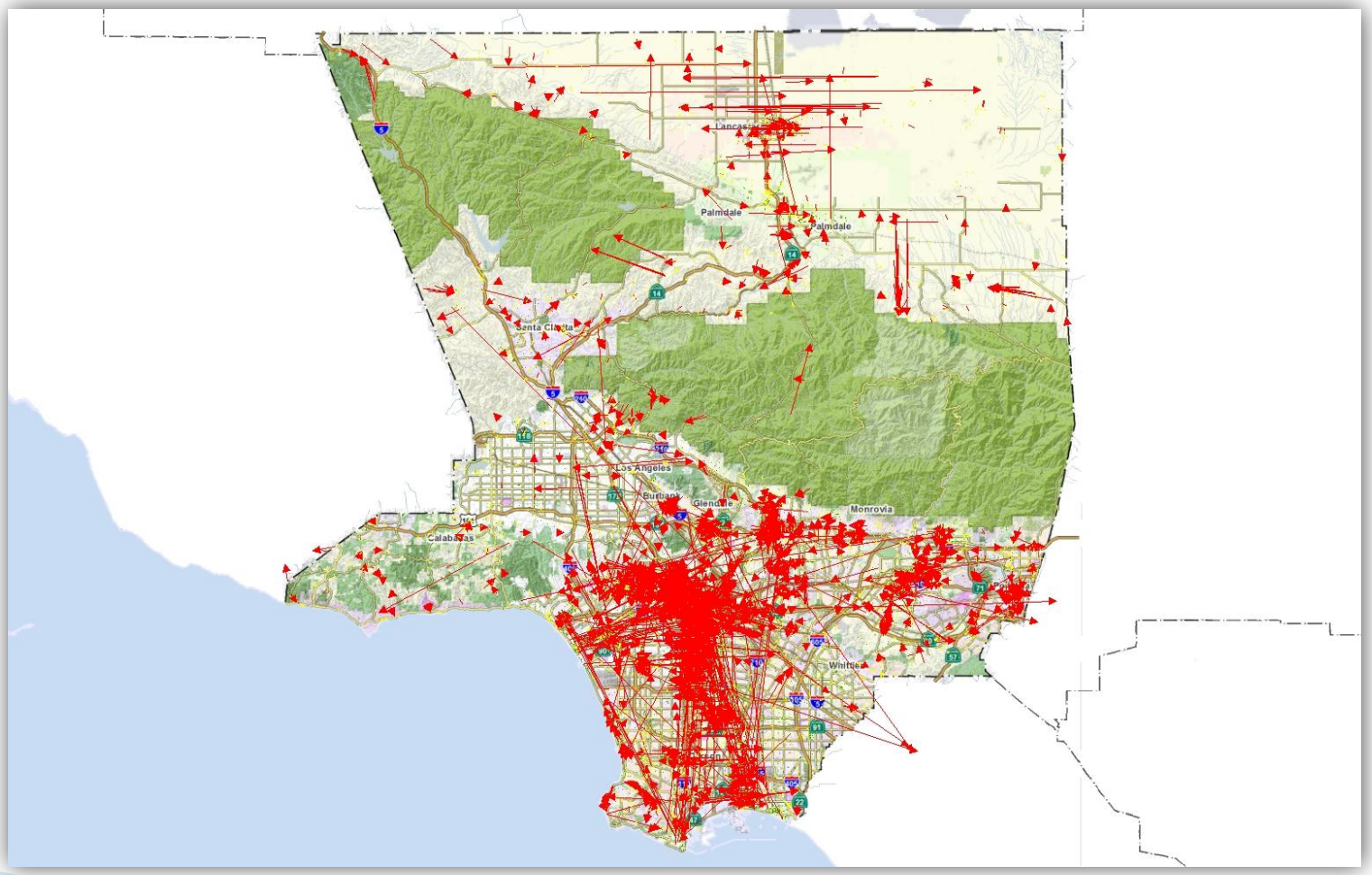


Alley Crossers

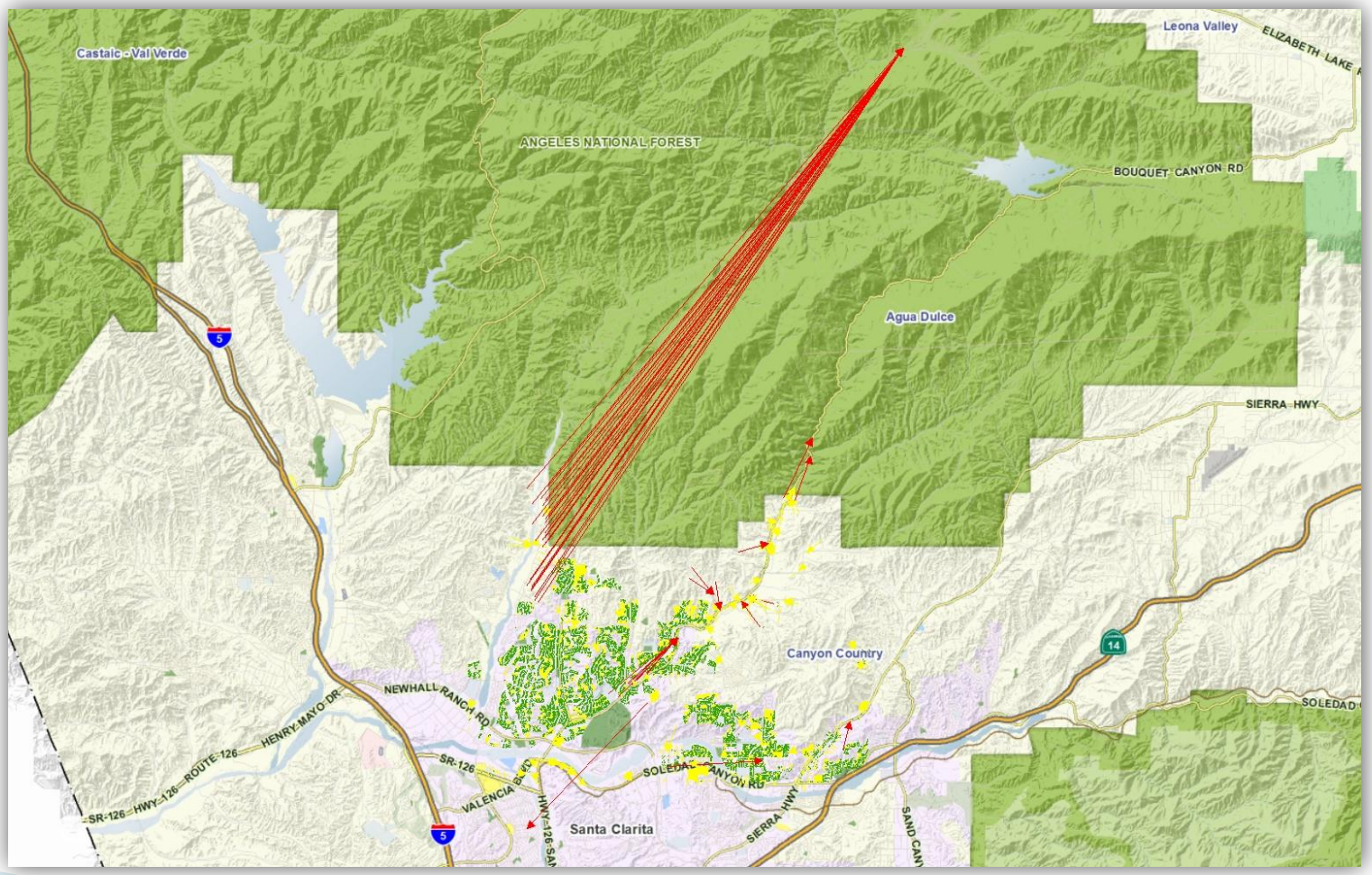
Crossers

- Street Segment error
- Off by just a few numbers
- Alleys can cause this, and street ranges that don’t end at “98” or “99” (e.g 398-399 Main Street).

Rockets


Lancaster
Palmdale
Palmdale
Santa Clarita
Los Angeles
Burbank
Glendale
Monrovia
Calabasas
Whittier

Rockets


Castaic - Val Verde
ANGELES NATIONAL FOREST
Leona Valley
ELIZABETH LAKE
BOUQUET CANYON RD
Agua Dulce
SIERRA HWY
Canyon Country
NEWHALL RANCH RD
HENRY MAYO DR
SR-126
HWY-126
ROUTE 126
VALENCIA BLVD
SOLEDAD CANYON RD
Santa Clarita
SOLEDAD
5
14

Flyovers/Rockets

Rockets

- Multiple error sources
  - Source data has zone error (should check) and can help fix source data.
  - Range error (a simple mis-type can cause this)
    - Range should be 28100 – 28198
    - It is 28100 to 38198 – catches many addresses
  - Incorrect zone information
  - Wrong directional information (It's East, should be West)
  - Matcher error (this is tough).

Recap: GIS Provides Unique Tools

- Visual patterns
  - Explosions
  - Implosion
  - Hops
  - Stretches
  - Spiders/Teepees
  - Magnets
  - Crossers
  - Rockets
- Color by line length
  - Short are green < 500 feet)
  - Medium are yellow (500 – 1,000 feet)
  - Long are red (> 1,000 feet)

Addresses in the future

- Where we are going next

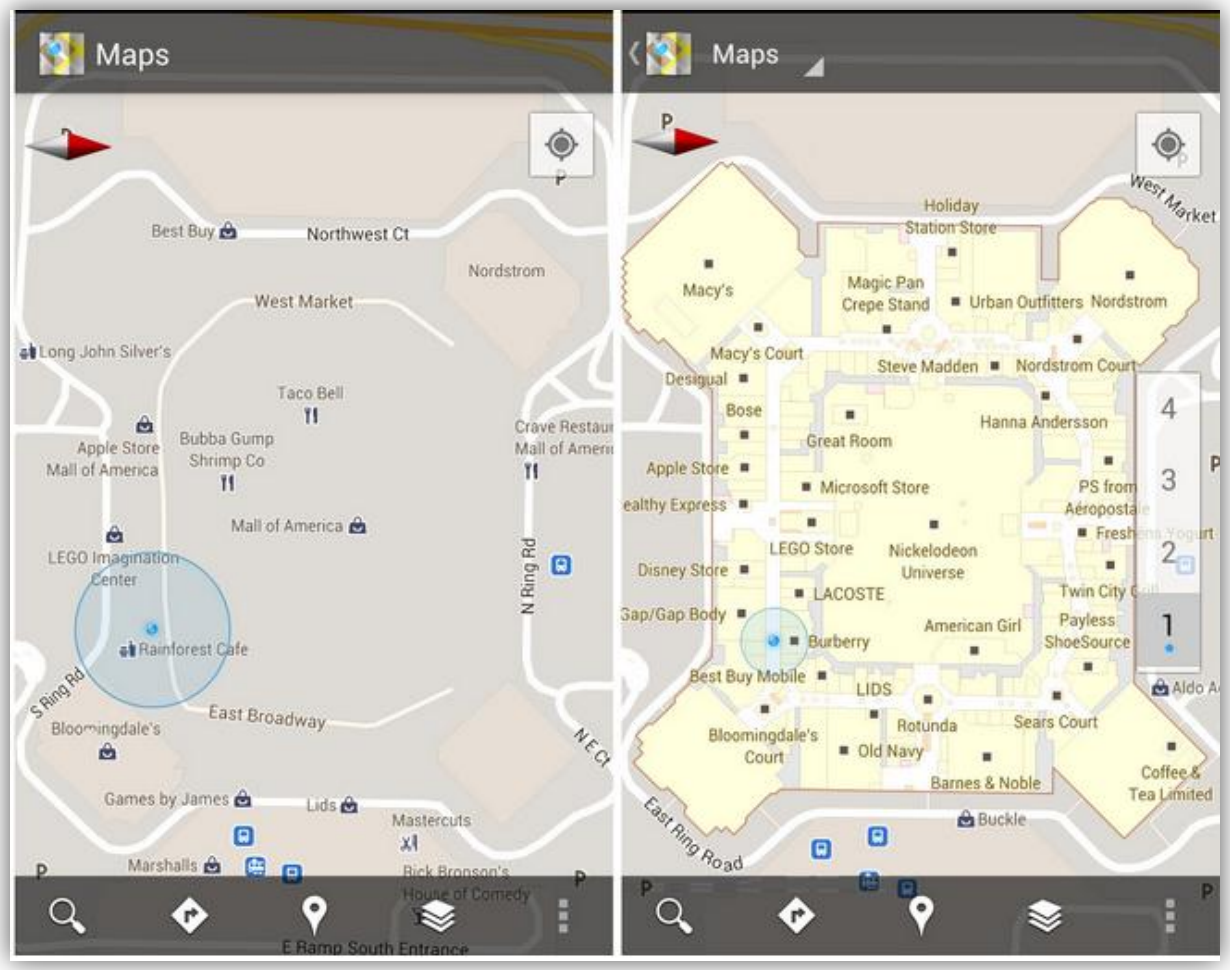
ESRI's new Matcher

- ESRI has released a new matching algorithm
  - New with ArcGIS 10
  - Single Line capabilities
  - Need to test it for speed, rejects, and false matches

The next 10 years

- Address Points
  - Point for each address and sub-address
  - Point for each suite, apt, internal building addresses
- Businesses
  - Handling quickly changing addresses
  - Malls, etc

Sample


Maps
Best Buy
Northwest Ct
Nordstrom
West Market
Long John Silver's
Taco Bell
Apple Store
Mall of America
Bubba Gump
Shrimp Co
Mall of America
LEGO Imagination
Center
Rainforest Cafe
N Ring Rd
S Ring Rd
East Broadway
Bloomingdale's
Games by James
Lids
Mastercuts
Marshalls
E Ramp South Entrance
Maps
Holiday
Station Store
West Market
Macy's
Magic Pan
Crepe Stand
Urban Outfitters
Nordstrom
Macy's Court
Steve Madden
Nordstrom Court
Desigual
Bose
Great Room
Hanna Andersson
Apple Store
Microsoft Store
LEGO Store
Nickelodeon
Universe
Disney Store
LACOSTE
Gap/Gap Body
Burberry
American Girl
Payless
ShoeSource
Best Buy Mobile
LIDS
Rotunda
Sears Court
Bloomingdale's
Court
Old Navy
Barnes & Noble
Coffee &
Tea Limited
Buckle
East Ring Road

Thank you.

Mark Greninger
mgreninger@cio.lacounty.gov

http://gis.lacounty.gov/egis
http://gis.lacounty.gov/dataportal

Questions?