Address Resolution from Survey 123 data

May 2022

Intro

- The aim: get the council property ids so information can be added to the property record

- Survey 123 already has address and location data – Right?

- So what is the problem – we can Geocode right?

Please Geocode these addresses:

Sample addresses:

	SurveyAddress
1	2,3,3a Hapua st Auckland
2	181 a Church st Auckland
3	357 Mt Roskill Road Auckland
4	see notes Biddick Rd Karaka
5	123lots1-4 St stephens Ave Parnell
6	Hibiscus coast hiway Orewa
7	421 Queen st Akl
8	12 Weona place
9	18a,b,c Alma Rd Milford
10	522 Great North Road Auckland
11	Smalley Glenbrook Auckland
12	
13	Lancom hou 29 Enfeild St Mt Eden
14	State hiway 16 waimauku

Uncorrected geocoded results

So what are the problems with the data

Address issues:
- No house number on survey
- House number doesn't exist (typo or unofficial address)
- Wrong road name
- No suburb (is this an issue?) – queen street samples
- Only house number (huh???)

What about GPS?
- Did you turn it on?
- Where were you when you submitted the survey
- Could you even get close enough to the property??

Examining the data

What sort of issues can we resolve in the data:

- Do some basic stuff (case changes, remove commas, #'s)
- Exact address match within property (Spatial Relator - 14%)
- Partial address within 200m (50%) (street number and name)

Can we geocode now?

- Sort of…
- Standard Geocode tool and options does not give us a property id

Geocoding?

How can we be confident that the match is correct?

- You can get 100% match if the address only has the street name!

- You can even get matches if there is no suburb:
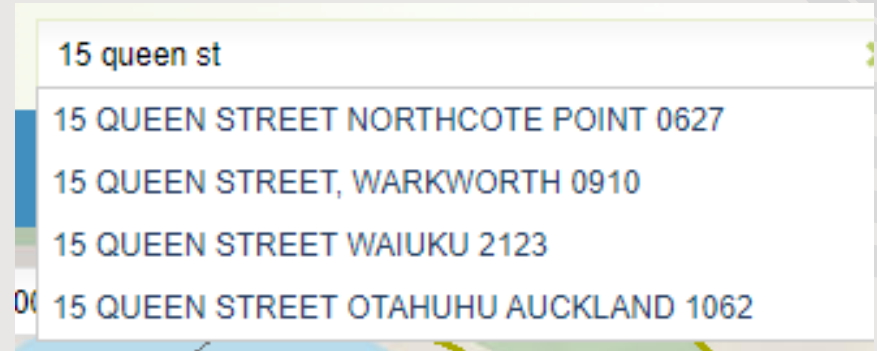
Capture some stuff first

- What happens with partial addresses (ie no suburb) where there can be multiple matches?

- Create an original X / Y coordinate so we can compare after geocoding

Geocode and check

- What happens with partial addresses (ie no suburb) where there can be multiple matches?

Problem?

We still have no property id
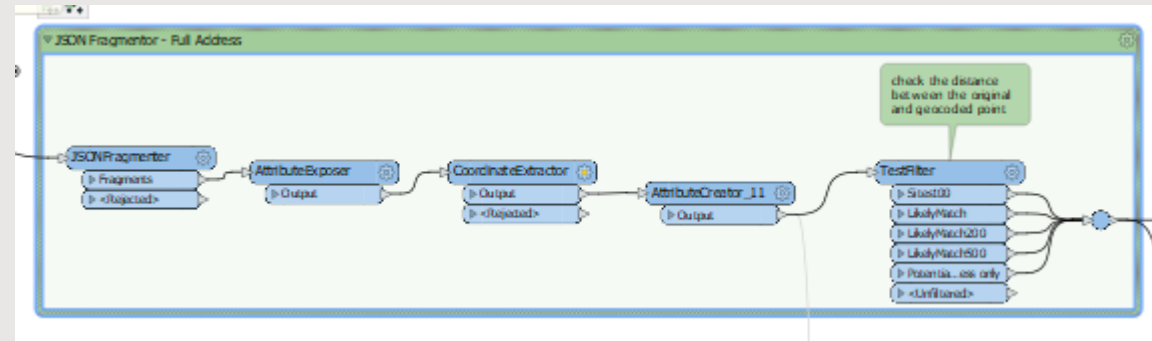
JSONFragmenter

Why?

- We can get the benefits of the geocoding

- Bonus – has property id for some matches

Issues

- Can still match to the wrong location if suburb is not supplied in the input data

- Address must have a house number or legal description

- Creates duplicates when there are parent / child properties with the same address – SEP (Someone Elses Problem…)
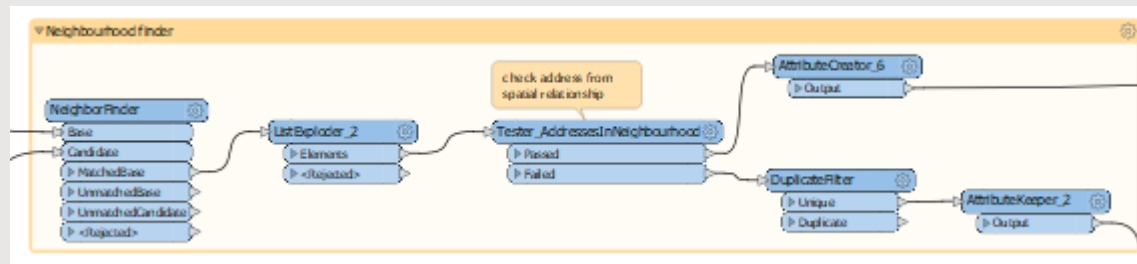
What next then?

How can I be confident that I have the correct property id?

- Expose the new coordinate from the geocoding service

- Look at the difference between the original coordinate vs the one from the geocoding service

- 100% match on address if the address has a number in it

- +/- 200m should be very good to give real confidence if no suburb is provided, and the match score is > 90

- +/- 500m should be good to give real confidence if no suburb is provided, and the match score is > 90
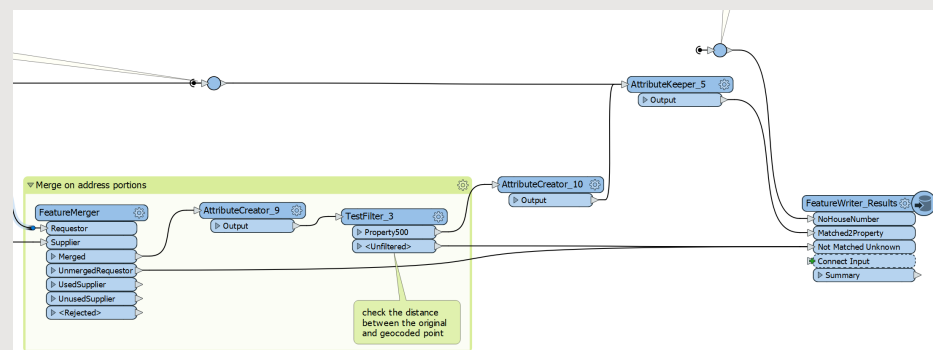
What's left?

Approximately 20% were not yet resolved

- Run a second JSONfragmentor using the partial field (no suburb)

- Recheck the differences in location

- Still only 250 more resolved – and some look like they should have been ok

- NeighbourhoodFinder – list potential addresses with similar number / road name within given distance

- FeatureMerger on address parts – surprised that this got some addresses that should have been resolved by the JSONFragmentor?

Outputs / Results

- Provide list of properties affected with confidence rating

- Provide list where no match was possible – no house number
  - Details might be on survey photos indicating address for cleanup

- Create a process to rename the survey PDF's to include the confirmed property id – allowing for multiple properties included in a single survey

Lessons:

- Fix the survey:
  - Minimise opportunities for stuffups!
  - Simplify address fields
  - Embed geocoder to allow selection of official addresses (and populate the property ID in the background - (avoid typos in the field)
  - Allow second location from point on an embedded map (good for flats or sites with multiple dwellings
  - Retain GPS
  - Extra field for reports that might be done for multiple properties
- Frequent reporting of unresolved addresses would allow gaps in training or data entry to be identified.

Questions?