Data Quality: utiliser le potentiel géographique

Vandy Berten – Smals Research



FOSS4G BELGIUM 2026

ceci n'est pas une conférence

Smals
ICT for society

Context

FOD Beleid & Ondersteuning
SPF Stratégie & Appui

- BeSt Address (by Bosa) is the authentic source for addresses in Belgium, gathering data from 3 regions
- Available as an API or as files (XML, CVS): https://opendata.bosa.be/

BeSt id	StreetFR	StreetNL	Num	Box	Zipcode	NIS	Geom
(…)/id/address/219307/1	Avenue Fonsny	Fonsnylaan	20	/	1060	21013	50.8358, 4.3384
(…)/id/address/219308/1	Avenue Fonsny	Fonsnylaan	20	2	1060	21013	50.8358, 4.3384
…	…		…	…	…	…	…

7,1 M

- Municipality code (NIS): includes one or several Zipcodes

Observations:

- From « classical Data Quality » point of view, BeSt is a very clean dataset
- Much better than RRN, KBO, …
- No « Av. Fonsny », « A FONSNY », « AVENUE FONSNY 20 », munic. vs sub-munic…

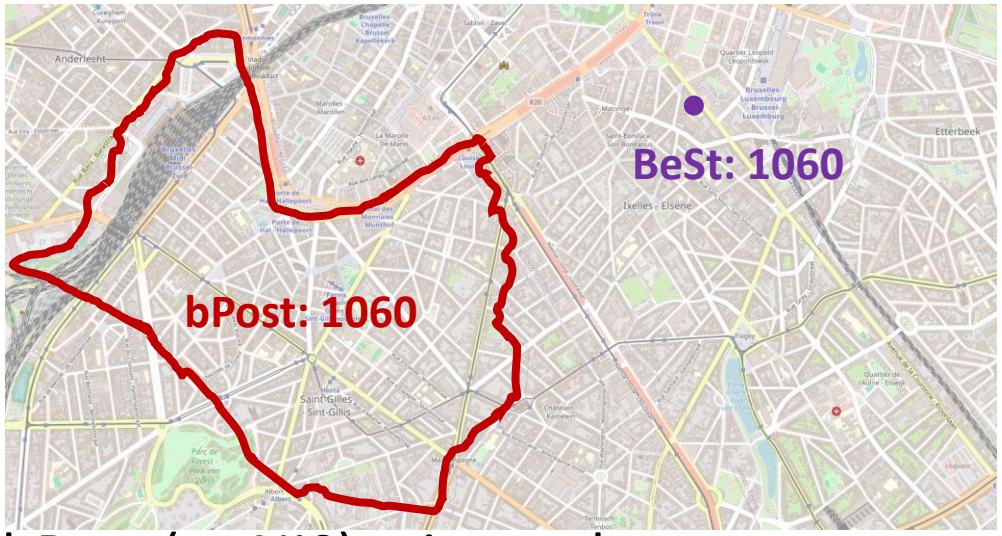
Research questions:

- Are there other quality issues?
- Can GIS help to find them?
- Address wrongly located?
- Street name wrongly spelled?

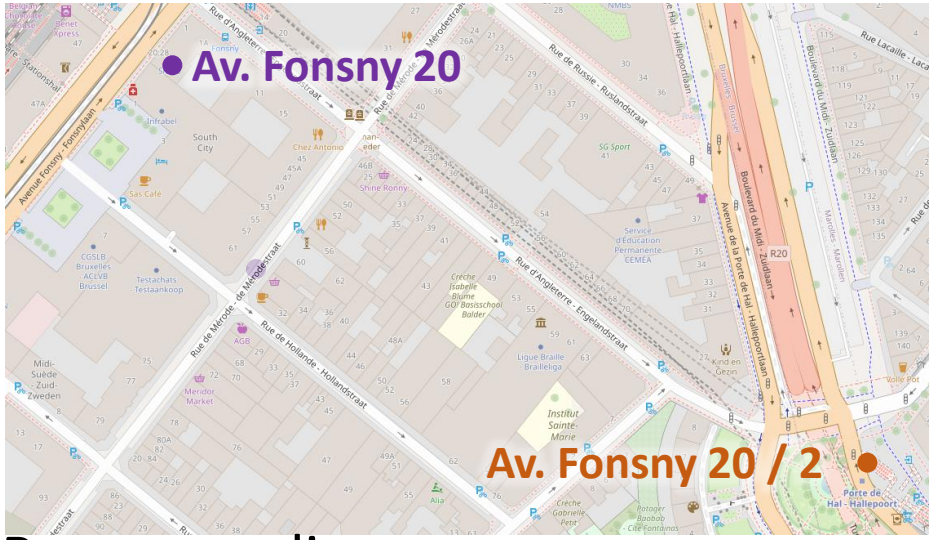
Spoiler alert: YES!

Smals ICT for society

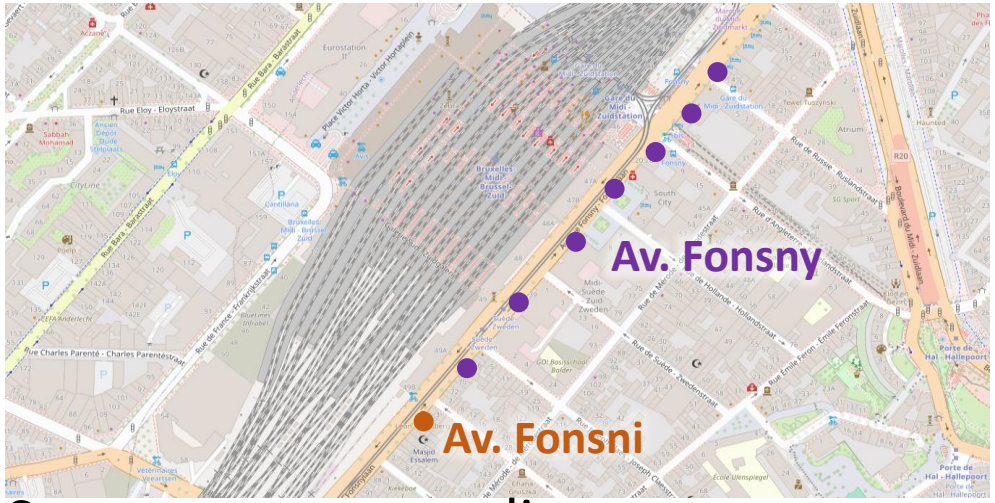
Anomaly typology



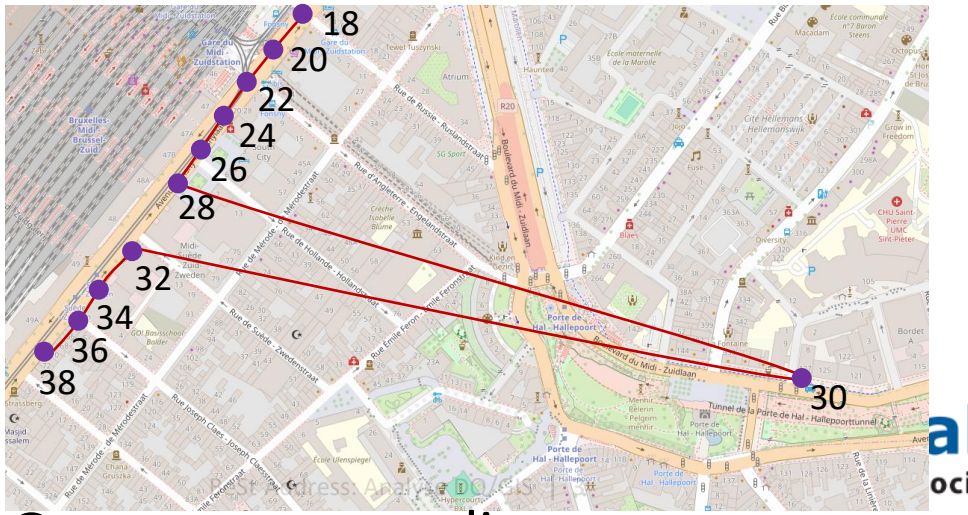
bPost (or NIS) mismatch



Box anomalies



Streetname anomalies



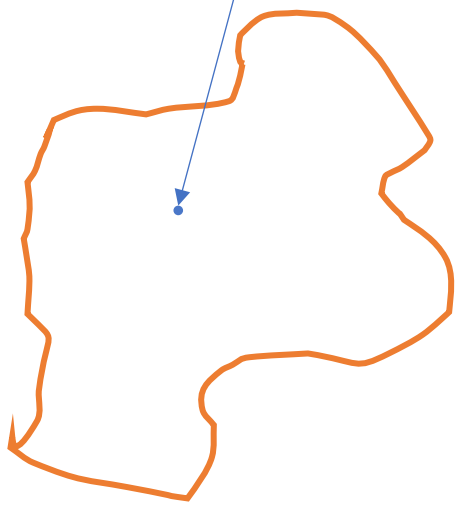
Geometry anomalies

Postal boundaries

BeSt →

bestId	Geom	PostalCode	…
548	472852.40, 6573959.524	5190	

bPost* →



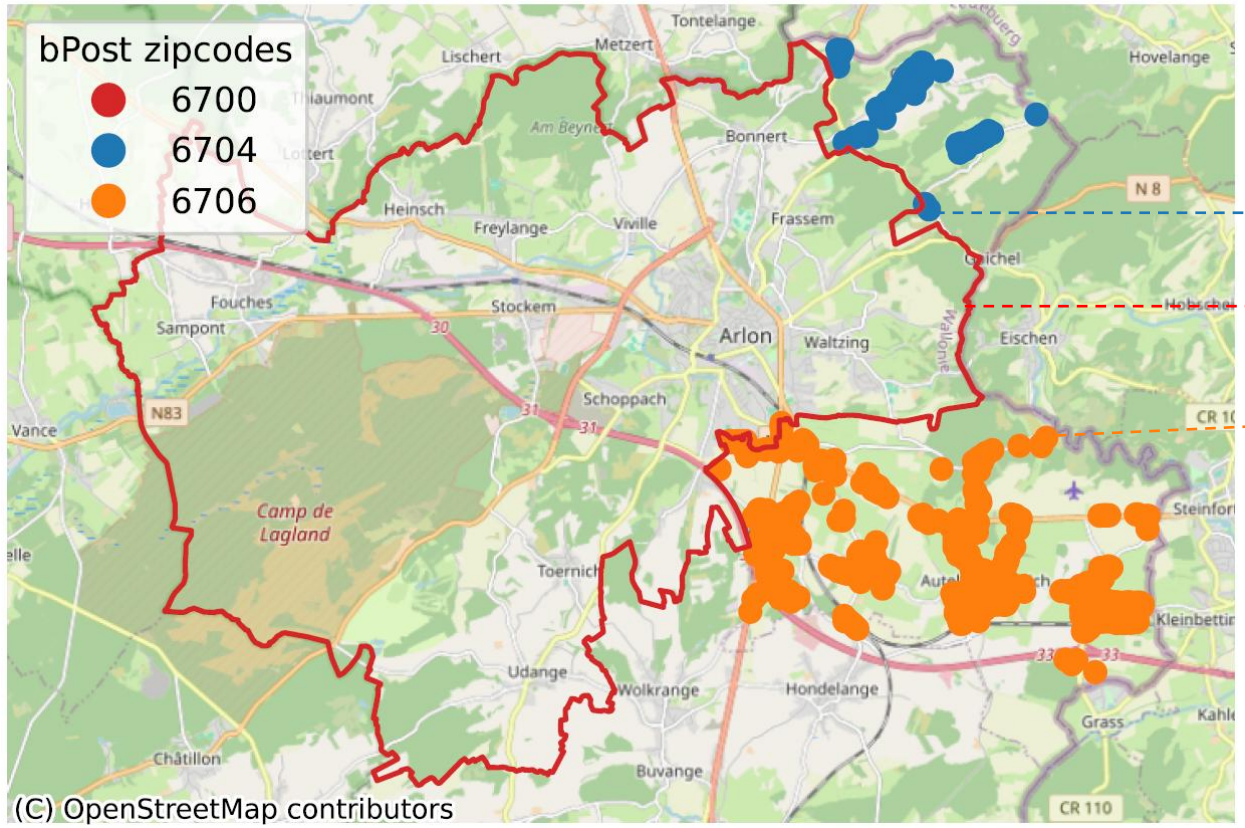
5190 ? **OK**
5180 ? **KO**

*https://www.geo.be/catalog/details/9738c7c0-5255-11ea-8895-34e12d0f0423?l=en

Smals ICT for society

Postal boundaries: local choice

BeSt zipcode / NIS: 6700 / 81001



bPost	BeSt
6704	6700
6700	
6706	

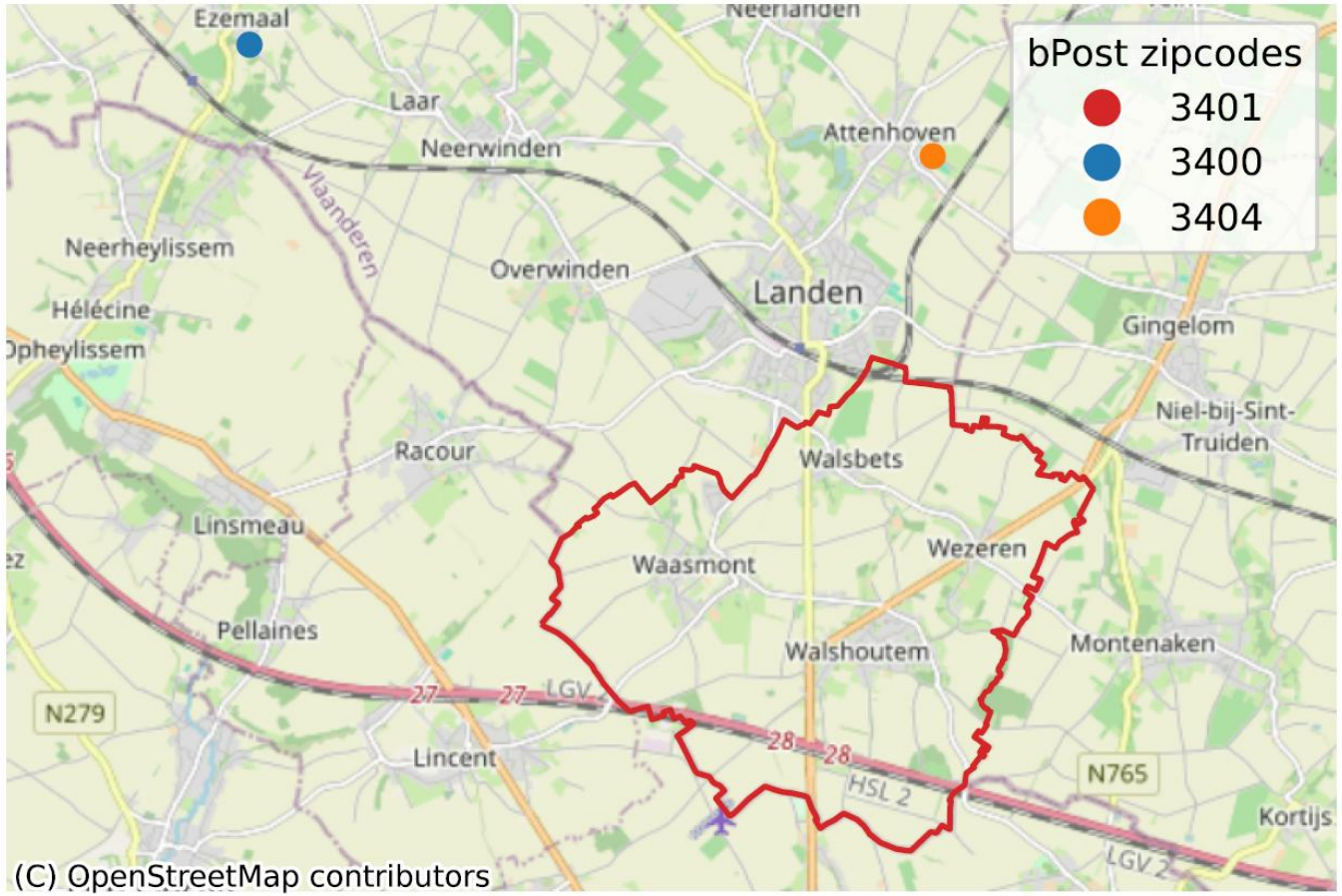
VLG : 9 600
WAL : 10 000
BRU: <50

- Clear choice of some municipalities to not apply bPost boundaries
- But (art 135, 21/03/1997 law): « postal codes (…) are the property of (…) bpost. They (…) can only be changed by bpost »

Smals ICT for society

Postal boundaries: mistakes?

BeSt zipcode / NIS: 3401 / 24059

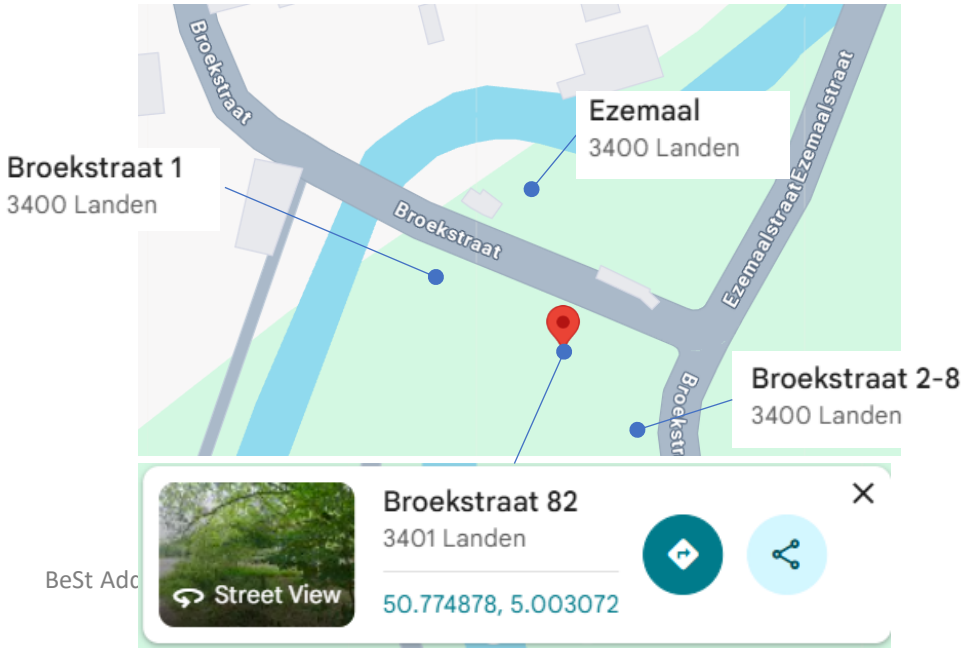


In BeSt: ● Broekstraat 82, 3401

Surrounded by « Broekstraat X, 3400 »

→ Zipcode likely wrong for this address, but correct coordinates

Idem in gMaps:



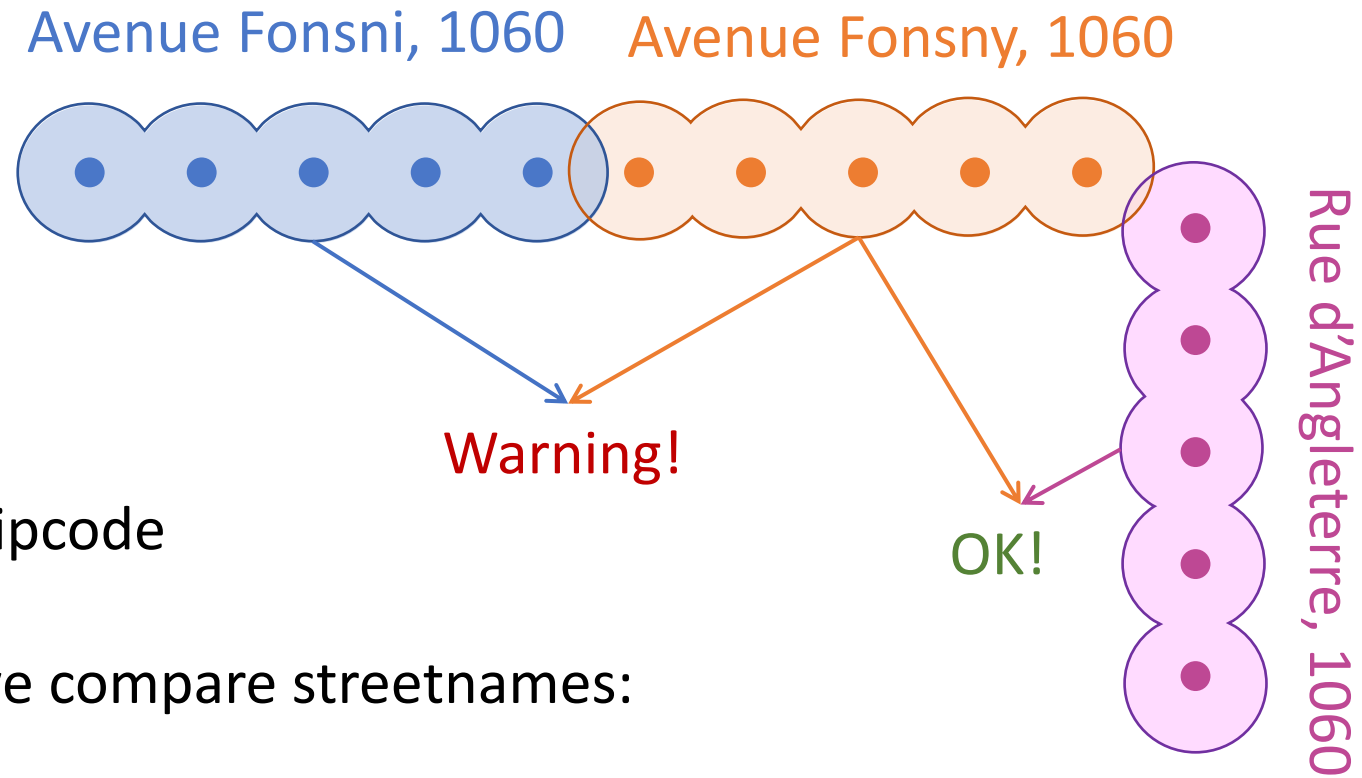
Municipality boundaries

- Same analysis with:
  - NIS codes (municipality_id)
  - Municipality boundaries (by StatBel)
- Same kind of errors (only « far errors »)



Smals
ICT for society

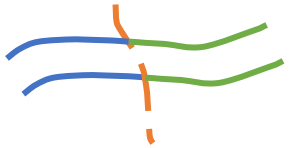
Streetname anomalies



- Group unique streetname-zipcode
- Buffer (100m)
- For overlapping polygons, we compare streetnames:
  - Close: Warning!
  - Not close: OK!
- Note: « close » is not that easy to compute!

Smals ICT for society

Name anomalies: streets crossing boundaries



Rue **de** Monténégro, 1060
Rue **du** Monténégro, 1190

Avenue de Tervu**e**ren, 1150
Avenue de Tervuren, 1160



Even: Rue Van **H**oegaerde, 1081
Odd: Rue Van**h**oegaerde, 1080

Rue Wal**c**kiers, 1030
Rue Walkiers, 1140





Rue Jacobs Fontaine, 1090
Rue Jacobs**-**Fontaine, 1020

Even: Rue Chapelle Baussart, 6001
Odd: Rue Chapelle B**e**aussart, 6032

Can municipalities choose their own street names? Yes, but*…

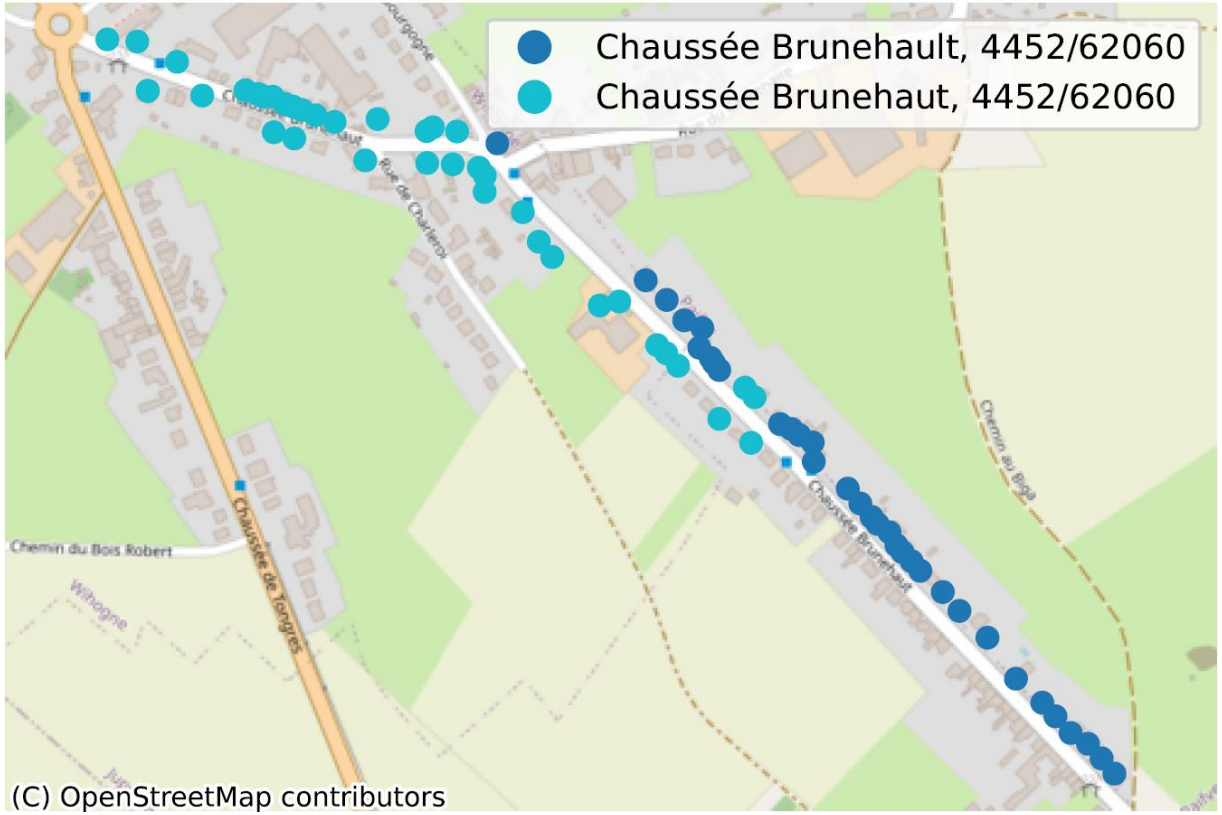
Where a road runs through the territory of **several municipalities** and retains the **same name**, those municipalities must ensure that the **spelling** of the street name is **identical** in all the concerned municipalities

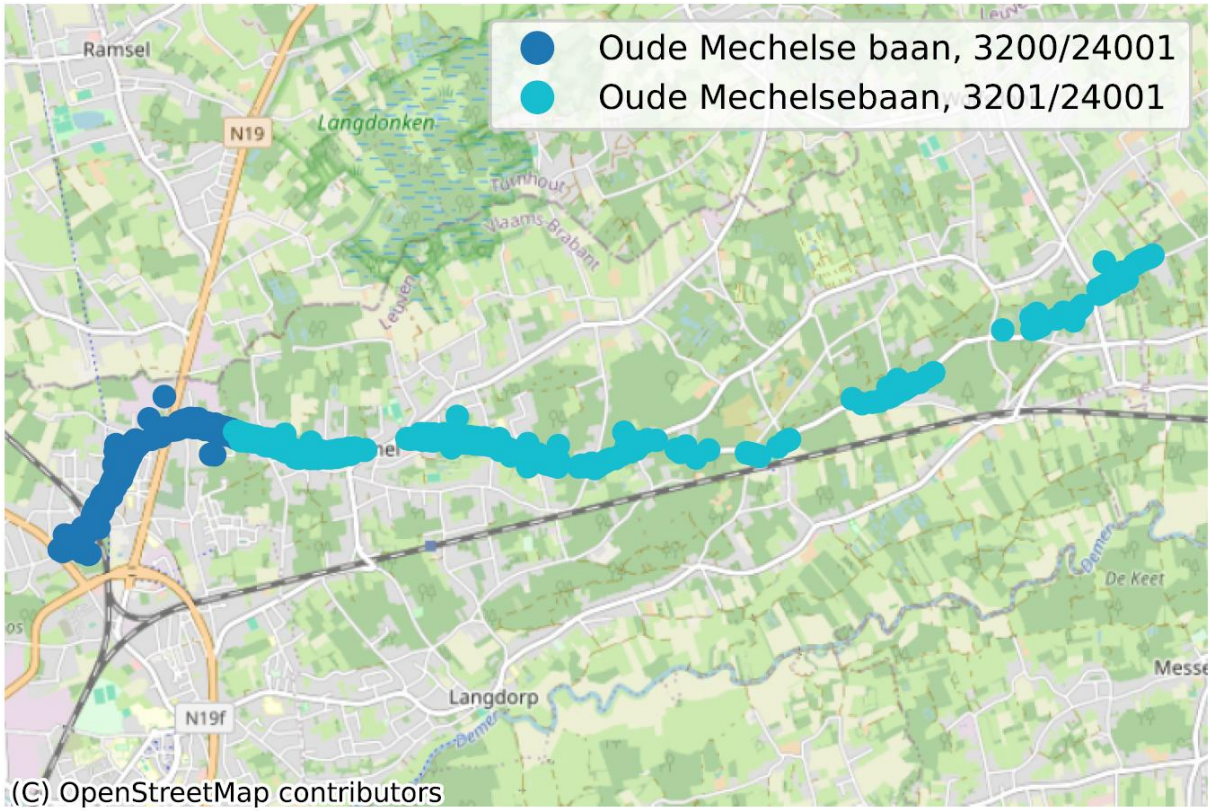
Each section of road must have the **same name on both sides**, even if those sides belong to different municipalities

* Guide en matière de constatation et attribution d'adresses (Directive SPF Interieur)

Smals ICT for society

Name anomalies: within a municipality

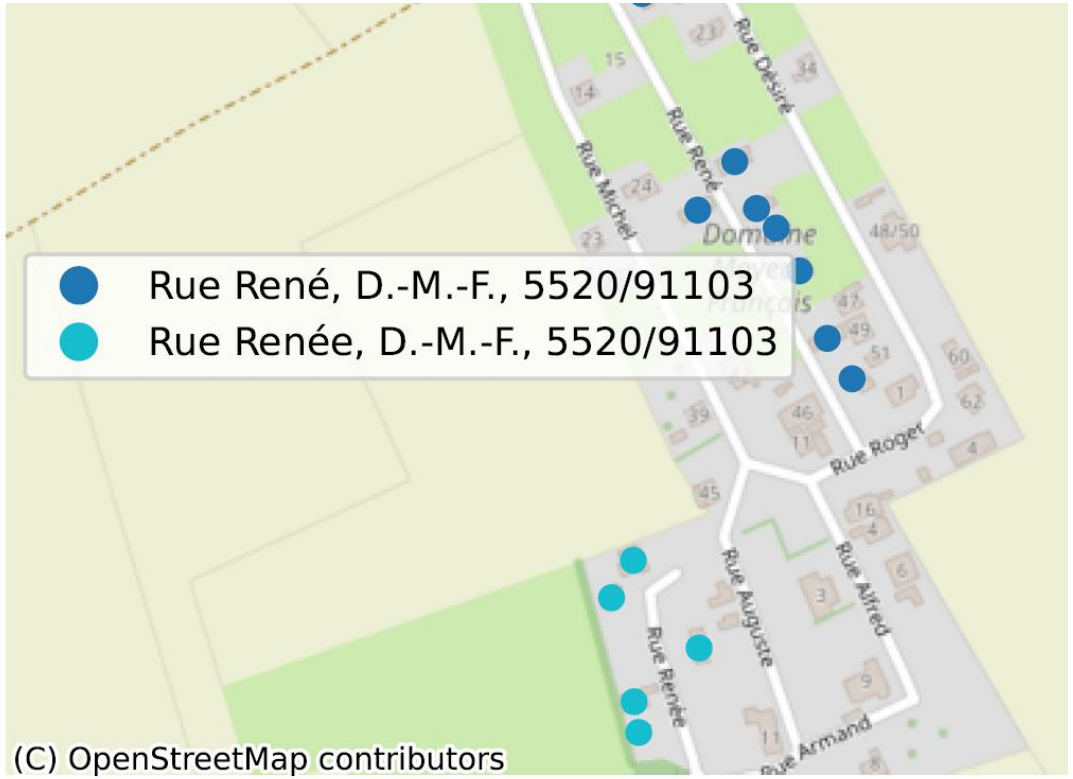


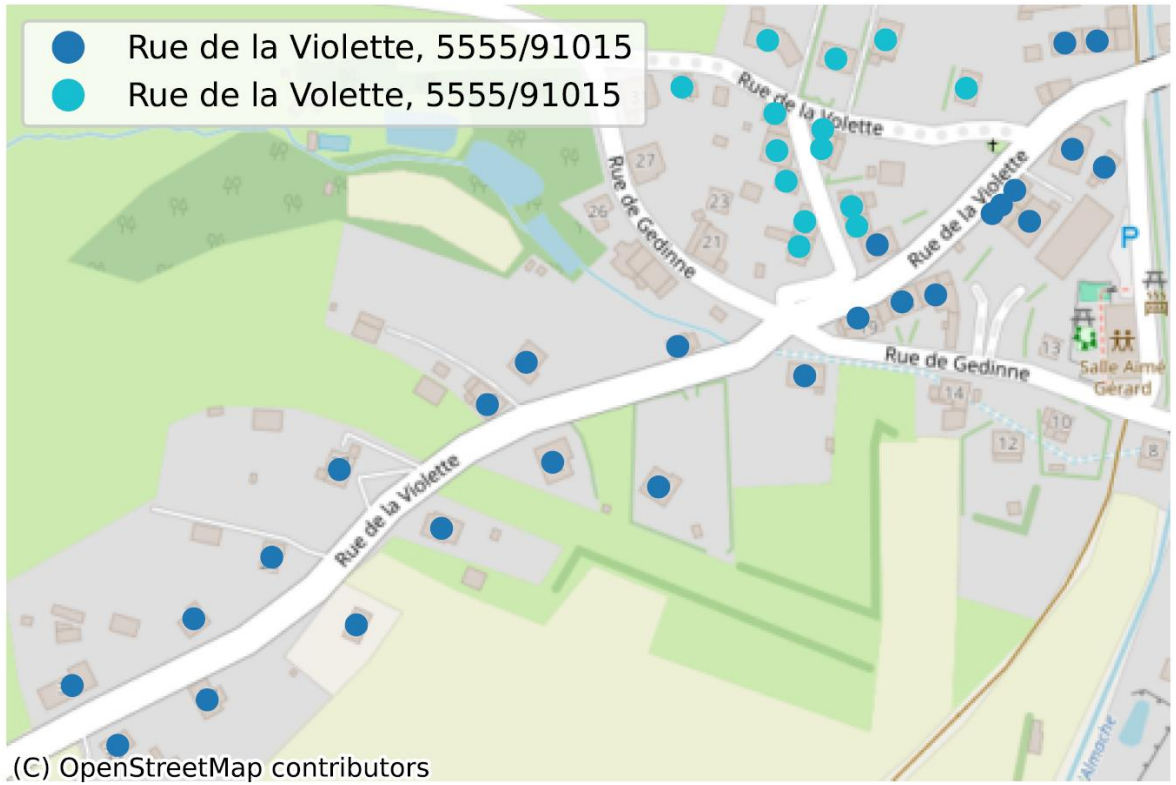


Ruelle [du] Pistolet, Rue de Sclaigneau[x], Rue des Chauf[f]ours, Rue de Forrière[s], Rue [d']Anixhe, Waalse[ ]baan…

Smals ICT for society

Name anomalies, confusing choices





Beukenlaan

Berkenlaan

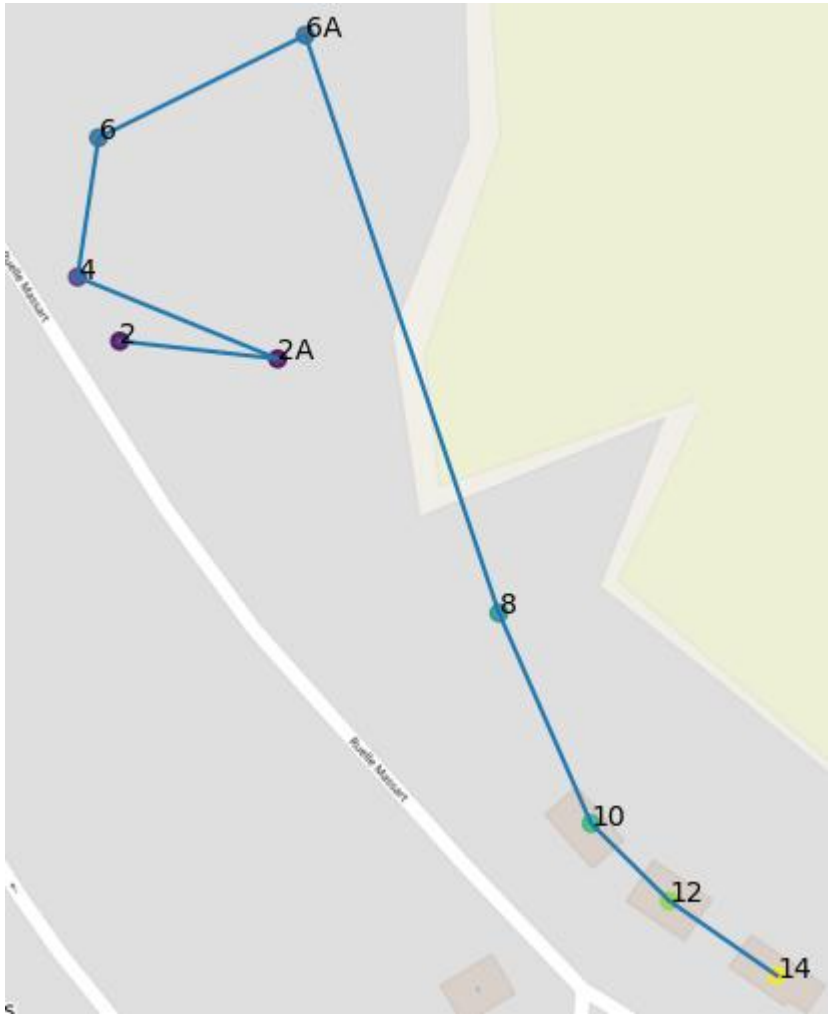
Burgstraße

Bergstraße

Smals ICT for society

Street geometry

- We consider a « line » with all addresses for
  - same street
  - same zipcode
  - same parity
  - order by housenumber
- We compute various metrics on those lines



Lowest number

Highest number

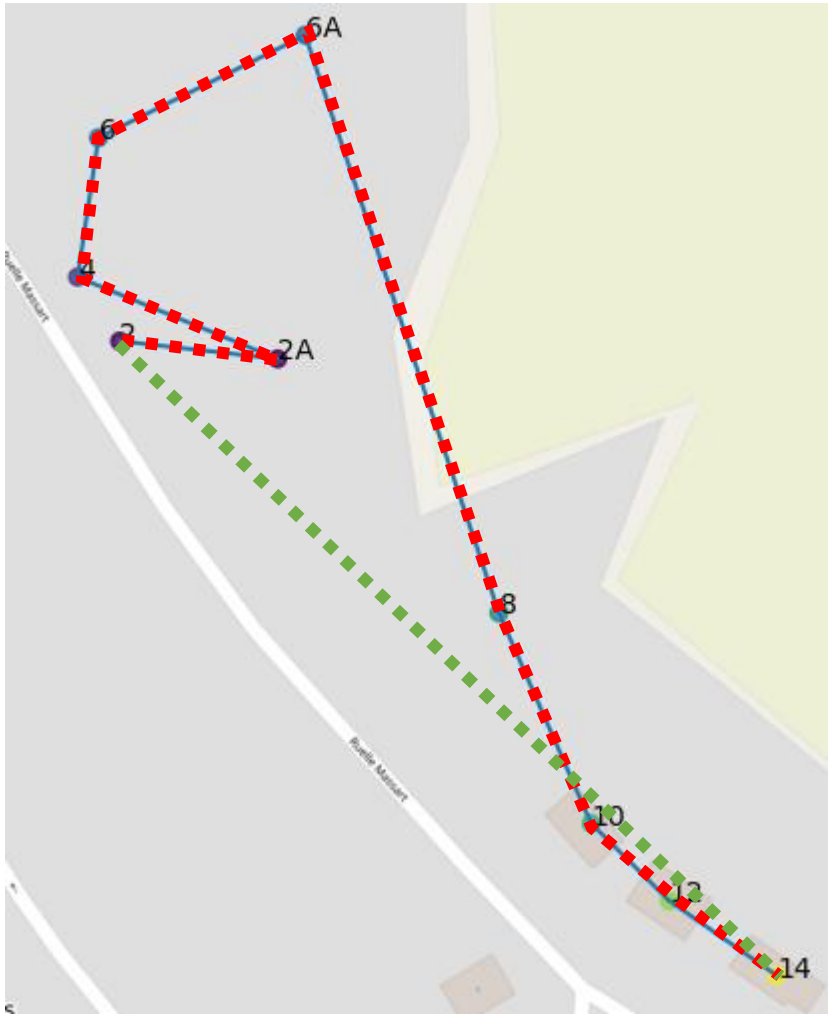
Smals ICT for society

Street geometry: sinuosity

- Sinuosity =

$$\frac{\text{Total length}}{\text{Straight distance between endpoints}}$$

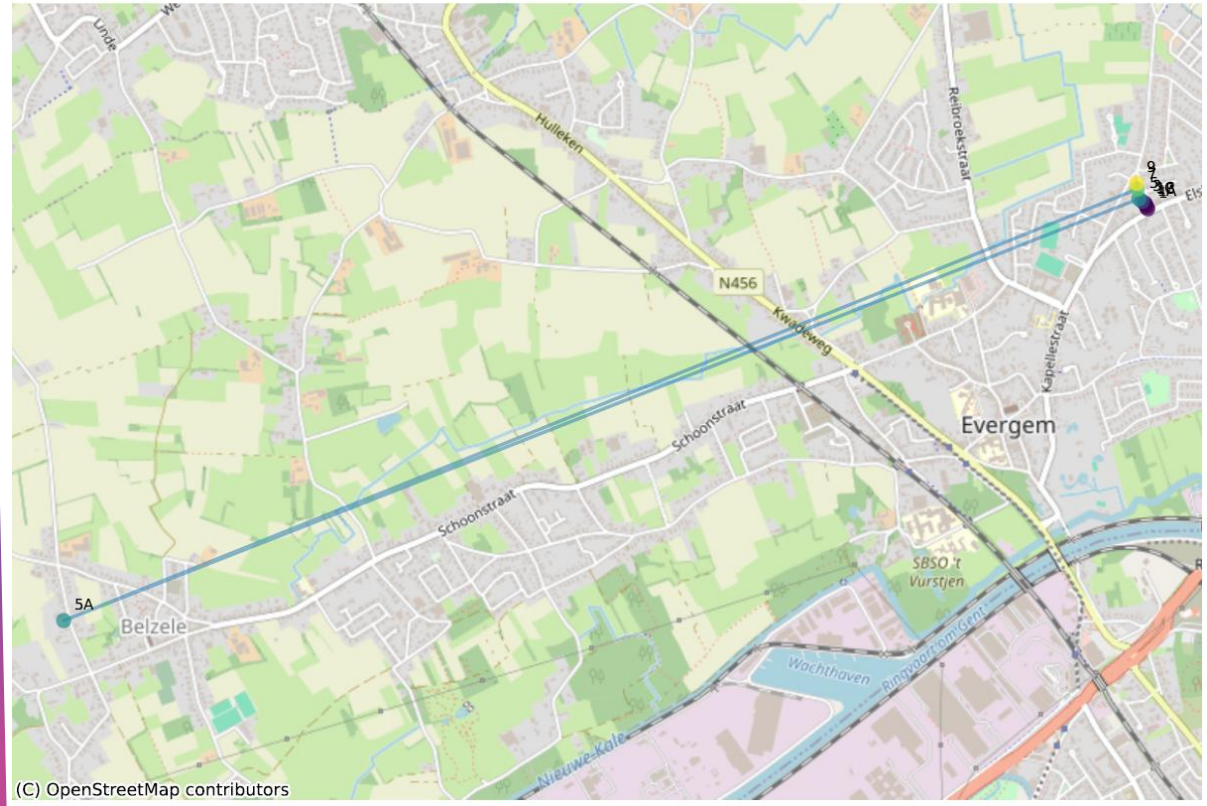
- High value usually denotes:
  - Wrongly placed addresses
  - « original patterns » (camping…)



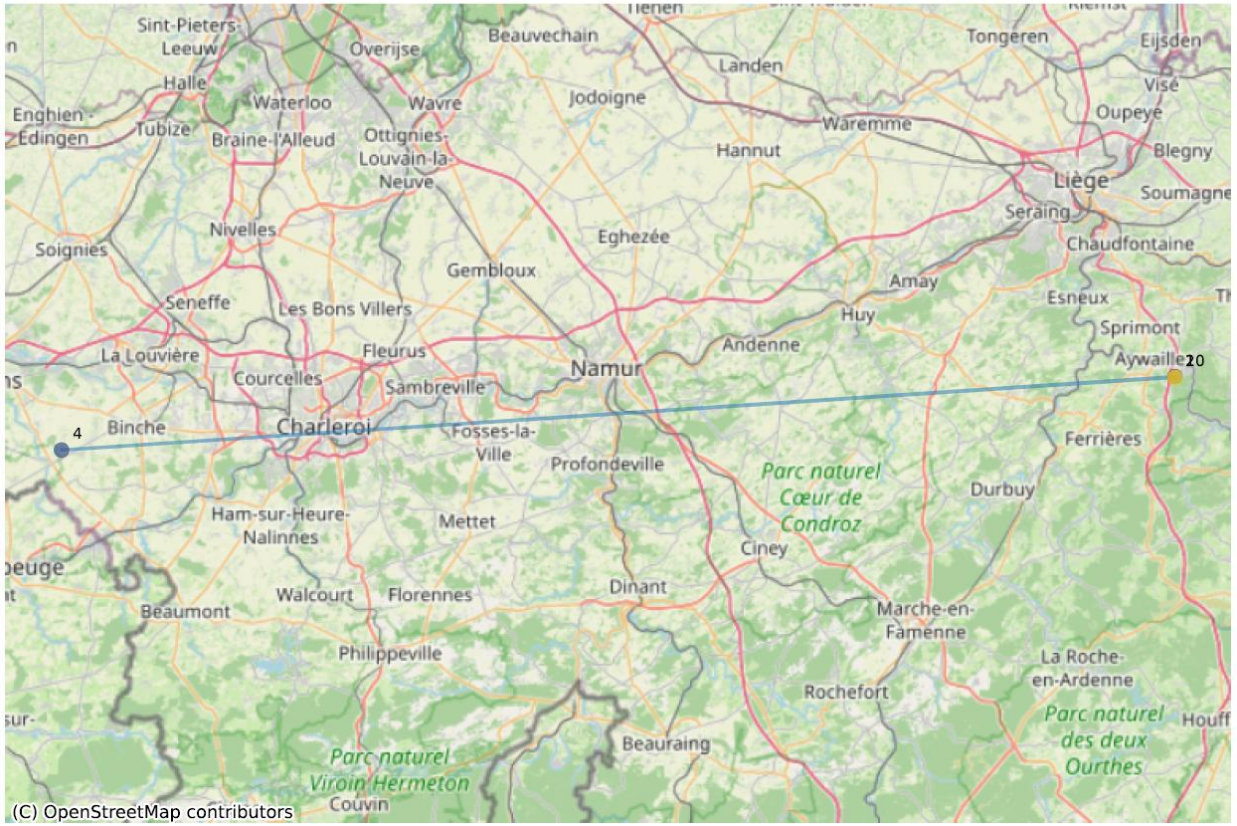
Smals ICT for society

High sinuosity: misplaced address

Kleine Koestraat - 9940/44019 - Ertvelde/EVERGEM/Kluizen/Sleidinge - Evergem - parity: 1
sinuosity: 83.23



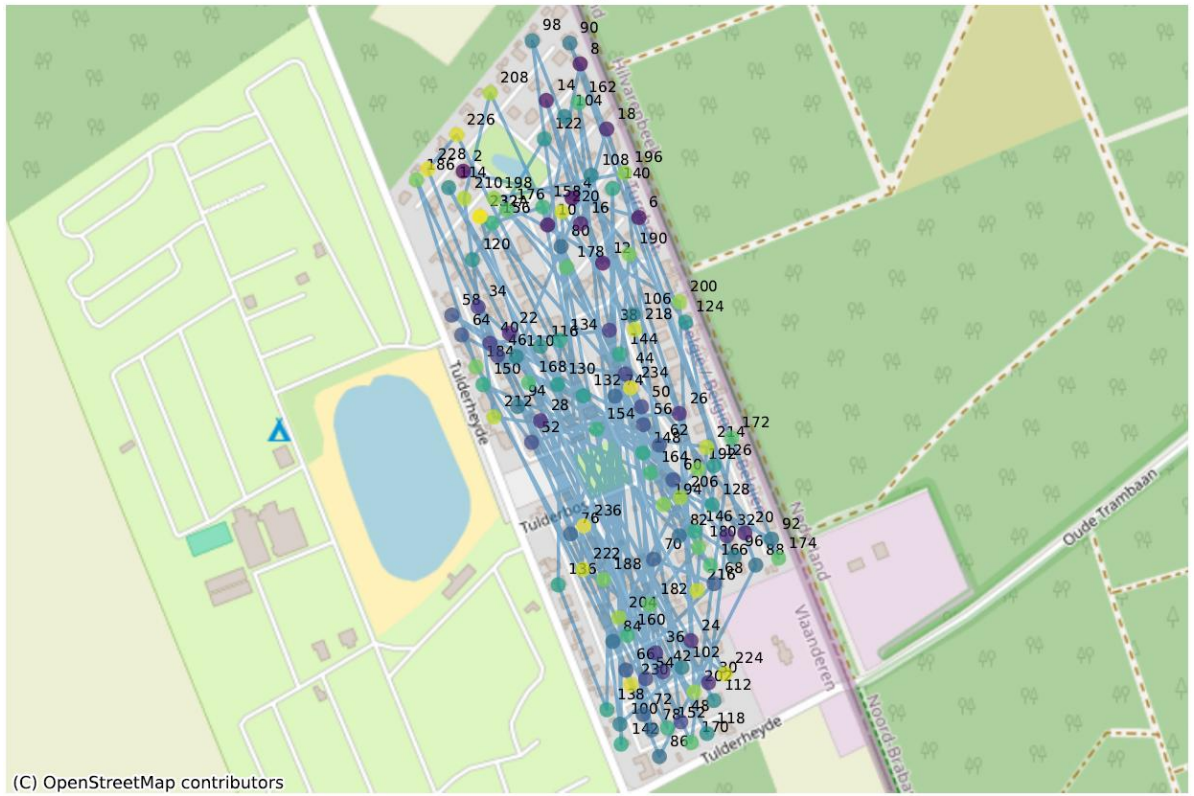
Allée des Brochets - 4920/62009 - Sougné-Remouchamps - Aywaille - parity: 0
sinuosity: 4100.81



Smals ICT for society

High sinuosity: non standard pattern

Tulderbos - 2382/13035 - Poppel - Ravels - parity: 0
sinuosity: 81.20

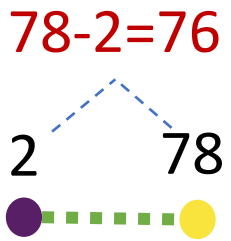


Boulevard de la Résistance - 1400/25072 - Nivelles - Nivelles - parity: 1
sinuosity: 126.21



Smals
ICT for society

Street geometry: delta_ratio

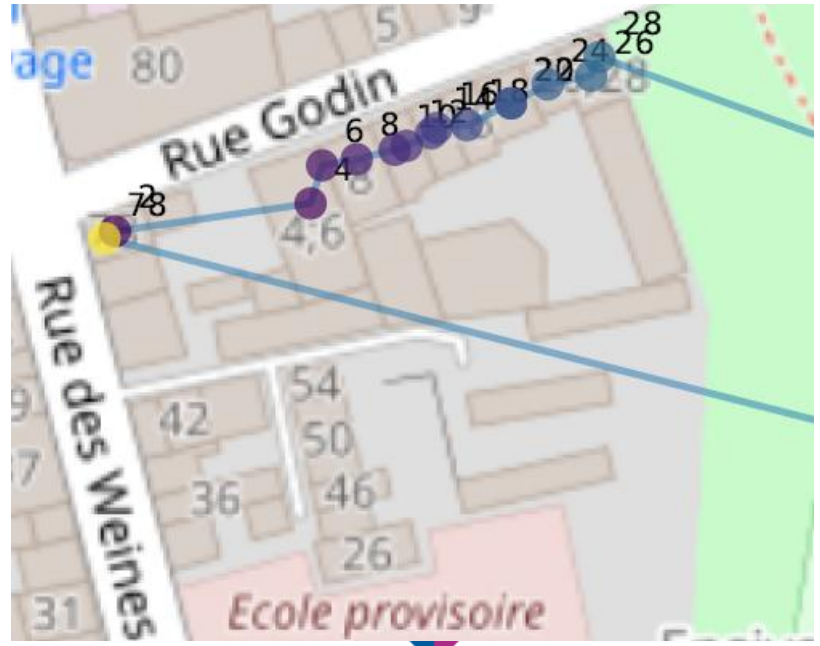




- ‘delta_ratio’ : max (for each couples of points) of

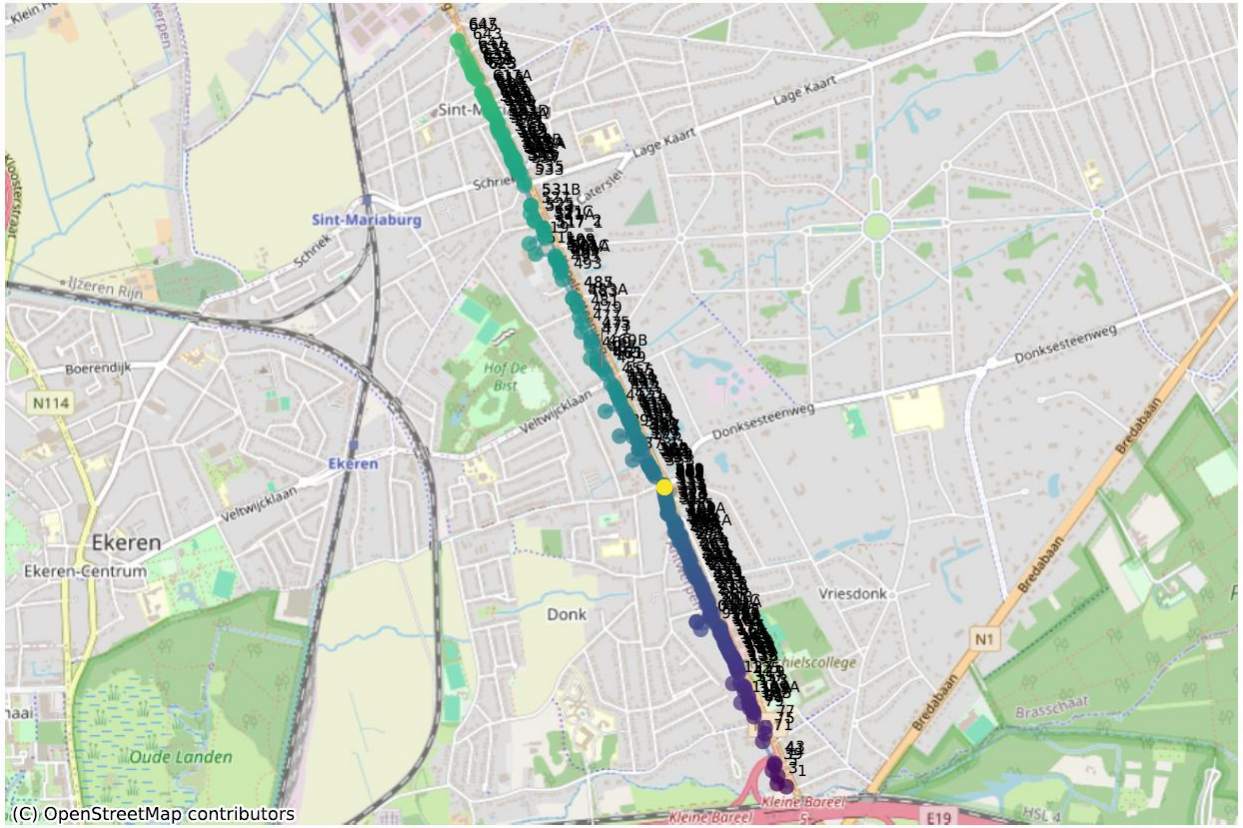
$$\frac{\text{Numerical distance between numbers}}{\text{Geographical distance}}$$

- Unit: Number/meter
- High value ➔ abnormally close numbers:
  - Circular streets
  - Wrong street on a corner
  - Wrong format (24001 instead of 24/001)

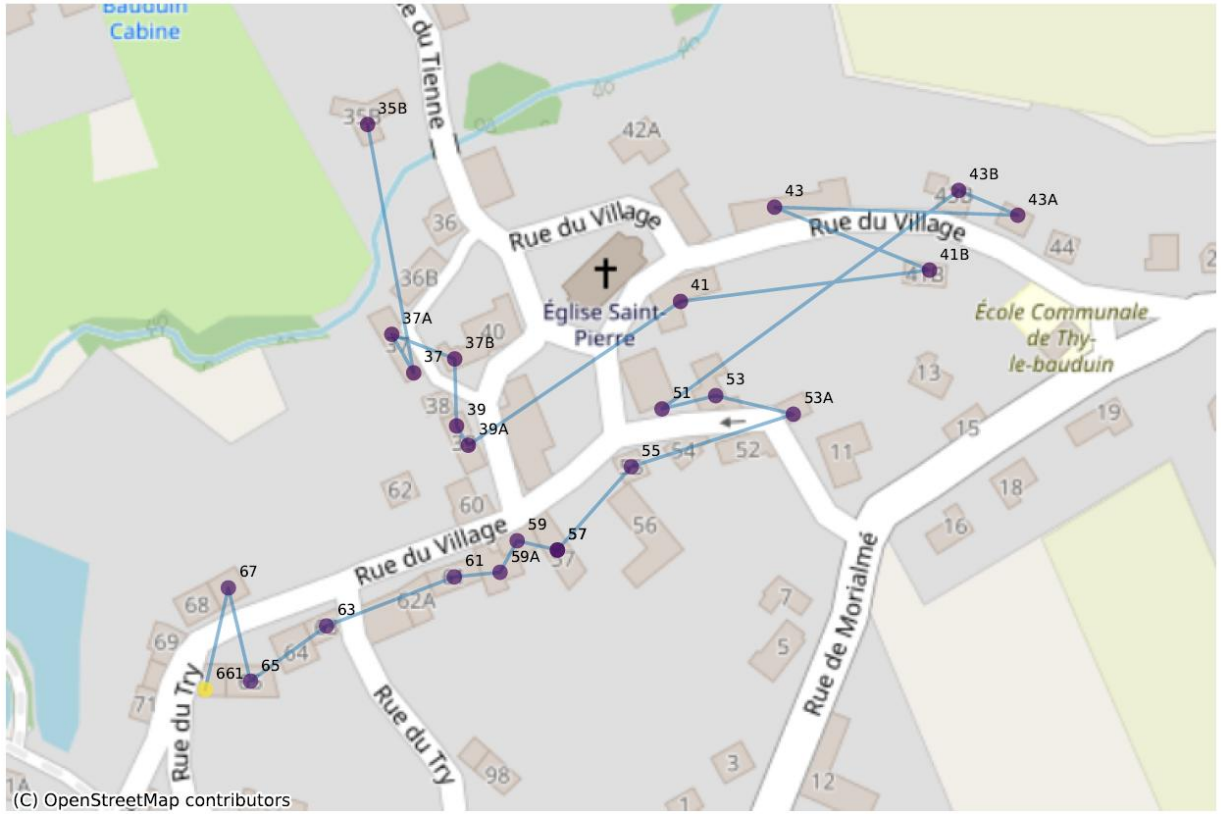


High « delta ratio »

Kapelsesteenweg - 2180/11002 - Ekeren - Anvers - parity: 1
delta_ratio: 221.78 (hn: 347 -> 929)



Rue du Village - 5621/93022 - Thy-le-Bauduin - Florennes - parity: 1
delta_ratio: 23.73 (hn: 65 -> 661)

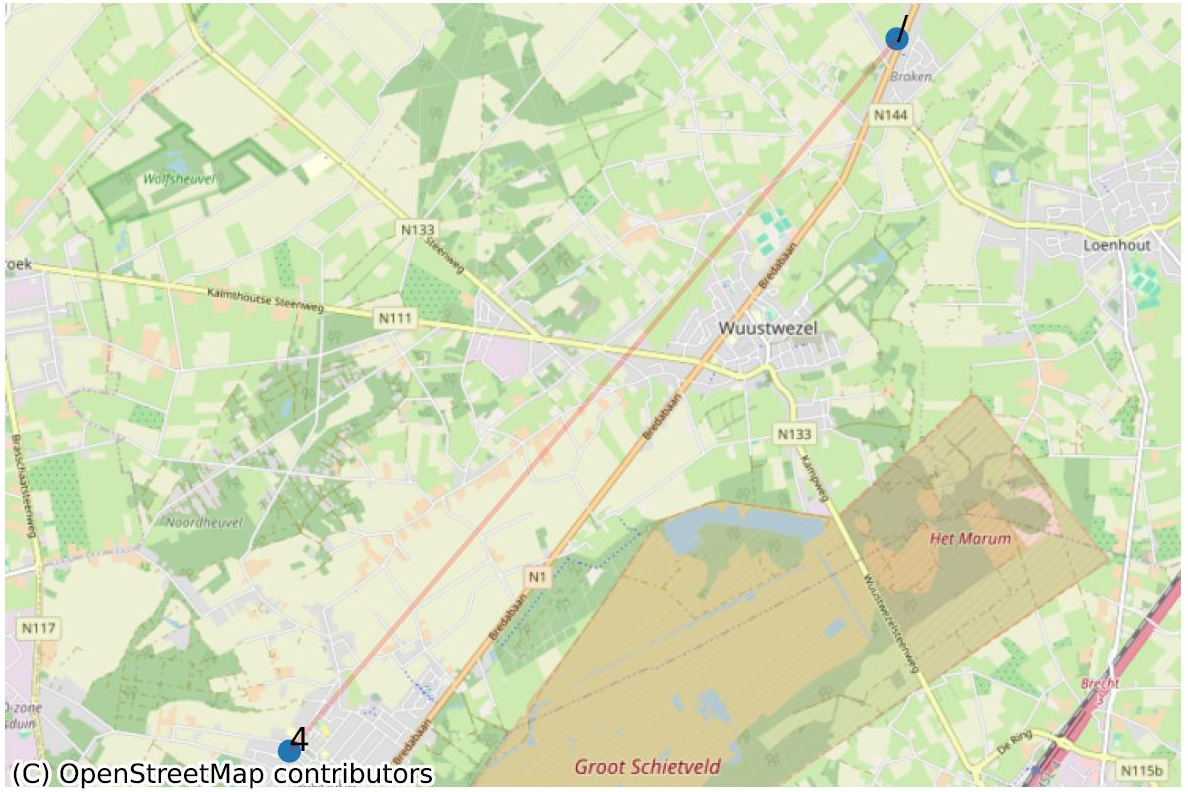


Smals ICT for society

Box anomalies

- Often, same house number boxes share same coordinates
- Not mandatory: in some residences, campings… all addresses share the same house number
- We detect situations where several addresses for the same number are far away
- Only in VLG (> 100 cases)

Bredabaan, 114, 2990/11053 Wuustwezel (13196 m x 0 m)



Smals ICT for society

Conclusion

Fast exploration	Quick analysis (~5 days) → we could go further if output was used
DQ x GIS	Original aspect: combining DQ / GIS technics
Clean baseline	Many other possibilities with « classical » DQ… but BeSt is very clean compared to other address sources (KBO, National register...)
Low count, **High impact**	Number of real mistakes low (<1000 vs 7.1M) … but large impact if emergency services are sent!
Open Stack	Only open-source tools (Python, geopandas, shapely) and open data (BeSt, bpost, Statbel)

- Python notebook: https://github.com/SmalsResearch/GISAnalytics/blob/main/BestAddressAnomalies.ipynb
- Blog post (FR/NL): https://www.smalsresearch.be/qualite-de-donnees-et-donnees-geographiques/ / https://www.smalsresearch.be/pourquoi-une-adresse-belge-cest-complique/

Smals ICT for society