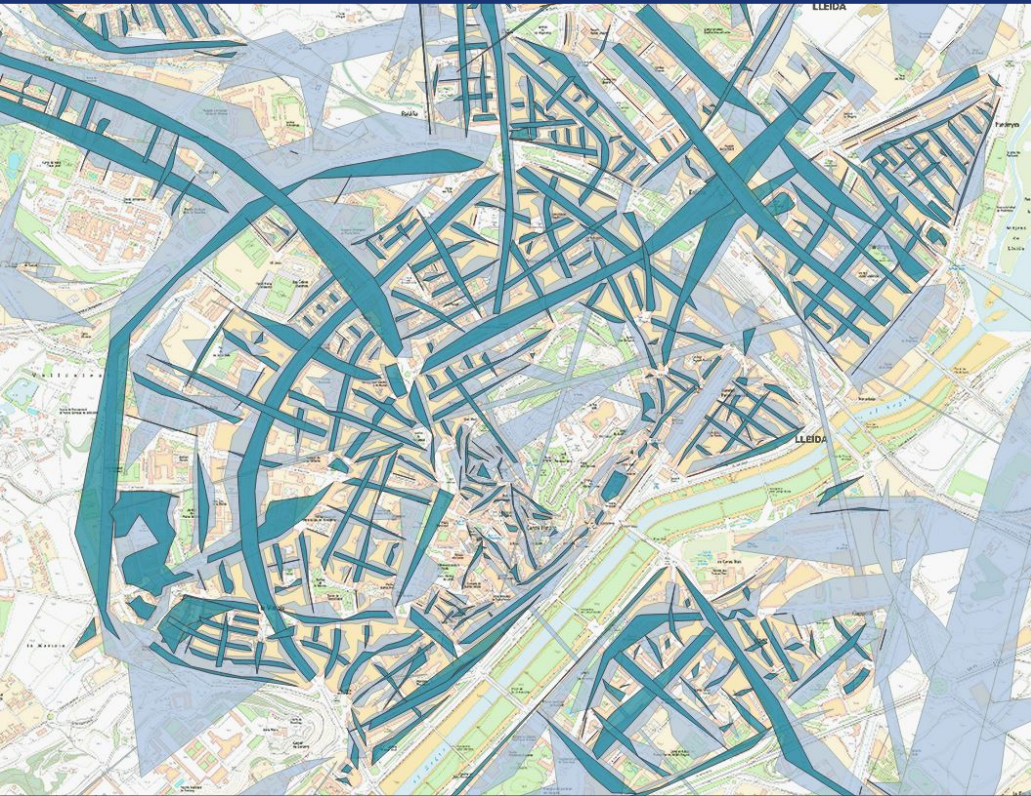




Generalitat de Catalunya
Institut d'Estadística de Catalunya

Implementation of the Statistical Territory Register

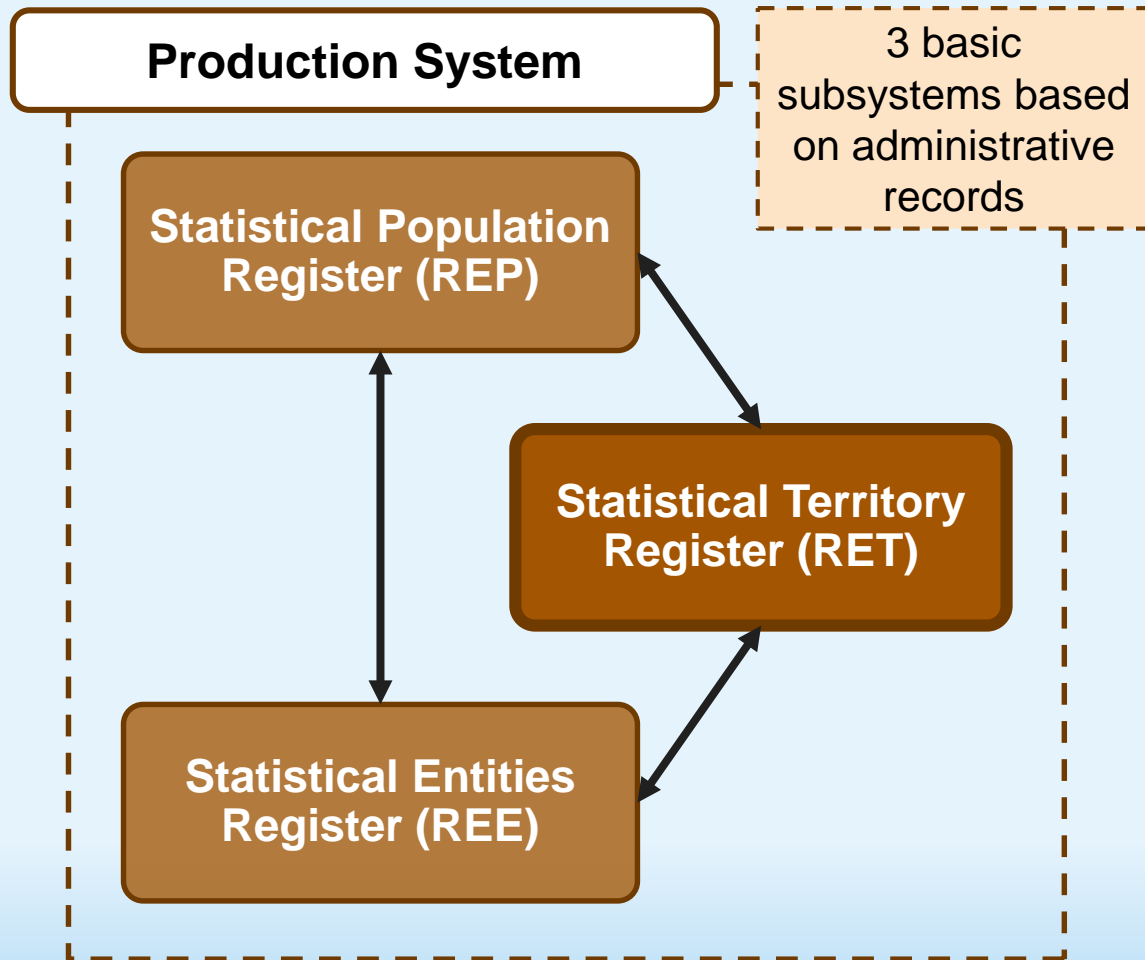


Cristina Rovira, Eduard Suñé, Anna Bernaus,
Daniel Ibáñez, Roser Condal, Mireia Farré

www.idescat.cat

- 1. RET: objectives and sources**
- 2. Geocoding**
- 3. Validation**
- 4. Quality control**
- 5. Conclusions**

1. RET: objectives and sources (I)



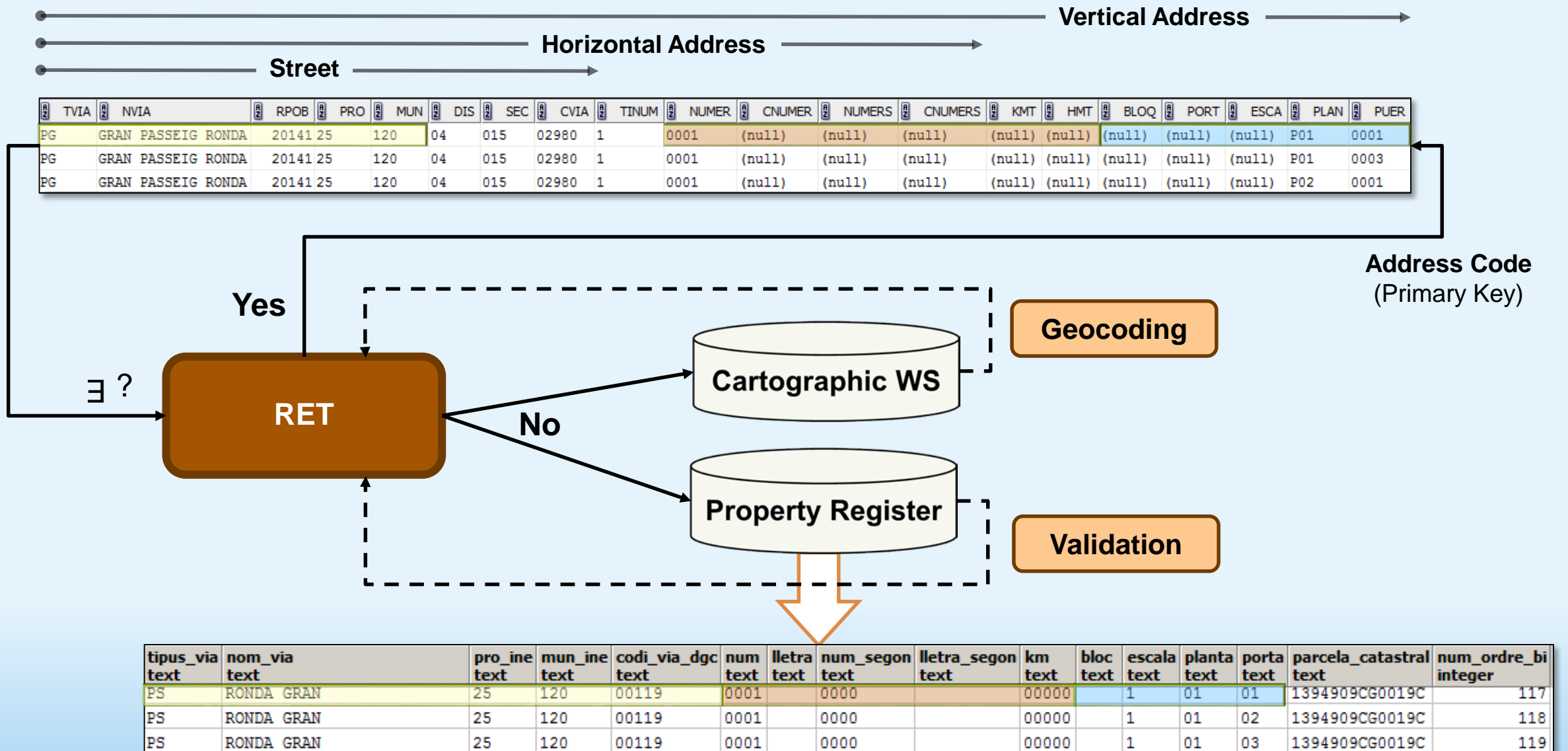
Objectives

- to geocode microdata
- to validate postal addresses

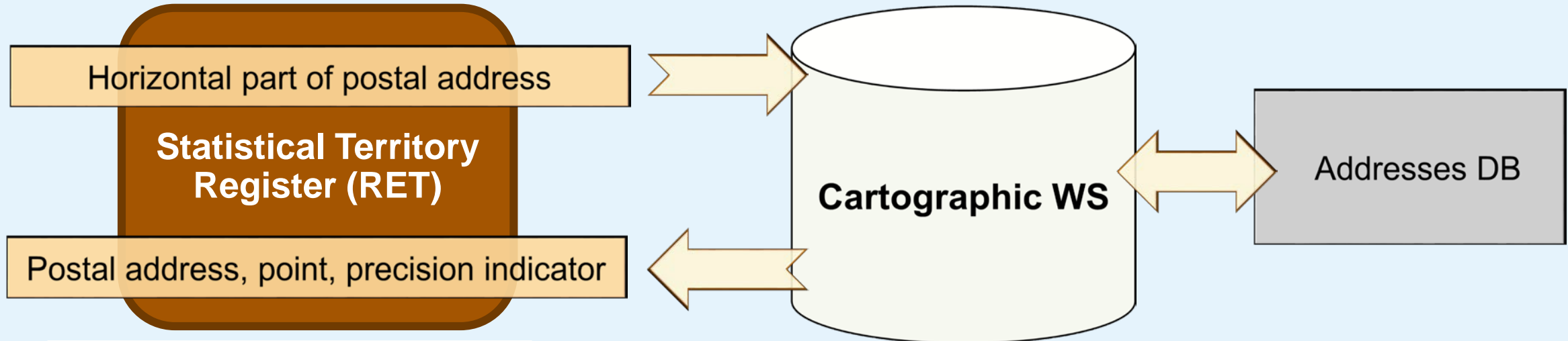
Sources

- Municipal database of addresses in Catalonia (only street numbers). Cartographic and Geological Institute of Catalonia (Cartographic WS)
- Plots of land and property. General Directorate of Land Registry (Property Register DB)

1. RET: objectives and sources (II)



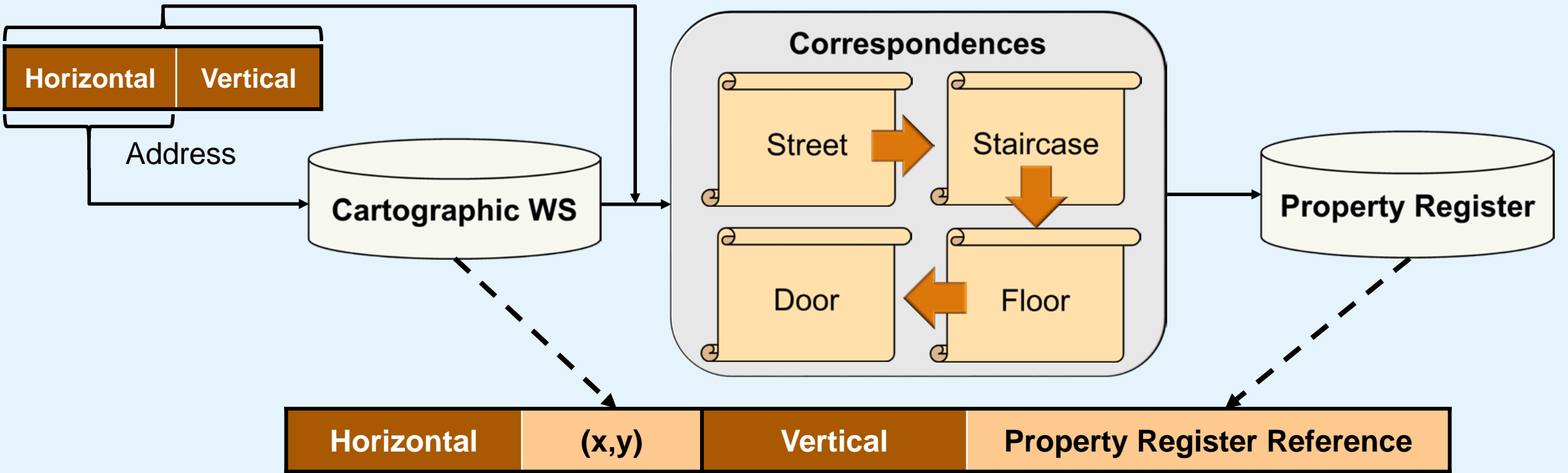
2. Geocoding



≈ 96% of addresses geocoded



3. Validation



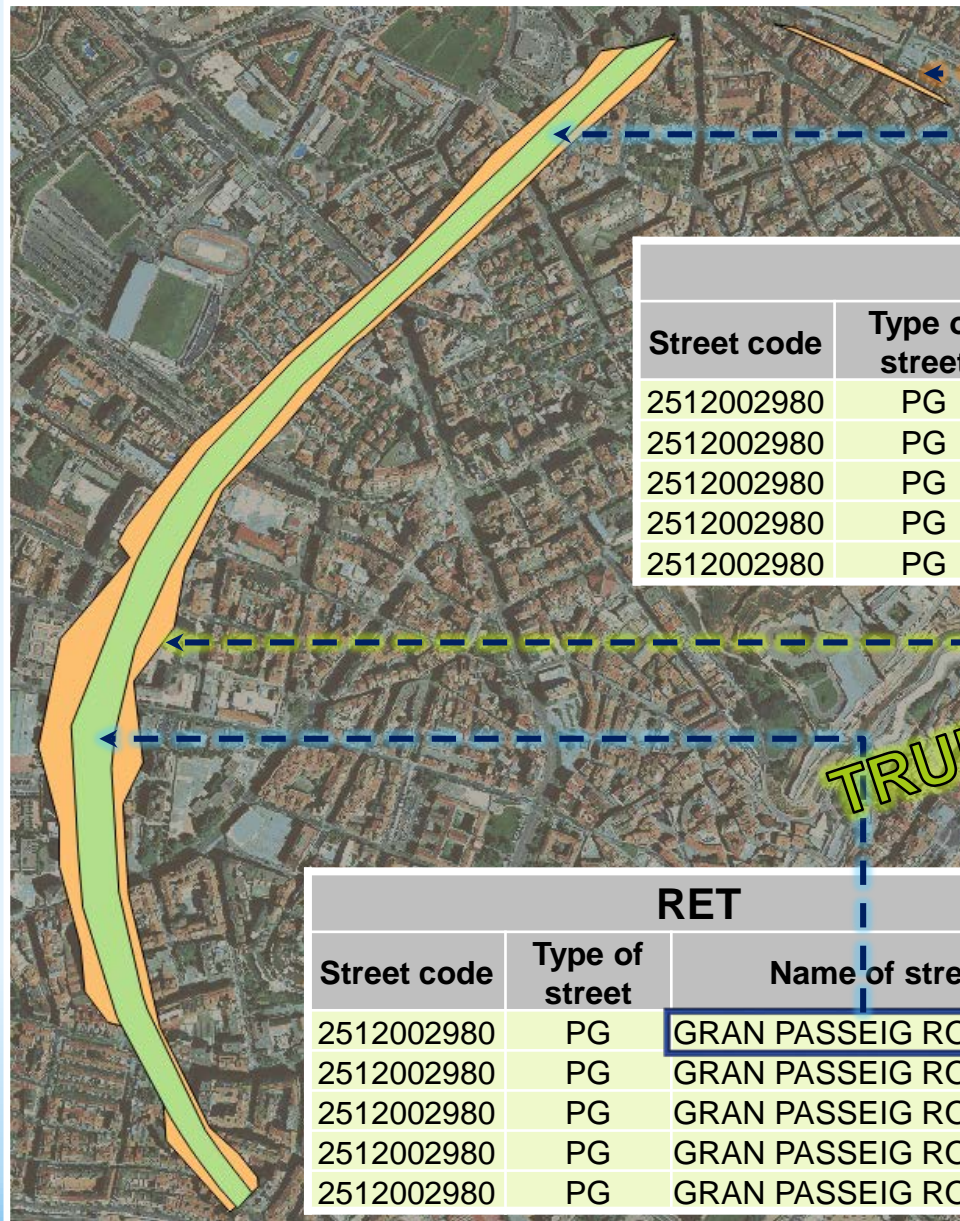
TVIA	NVIA	RPOB	PRO	MUN	DIS	SEC	CVIA	TINUM	NUMER	CNUMER	NUMERS	CNUMERS	KMT	HMT	BLOQ	PORT	ESCA	PLAN	PUER
PG	GRAN PASSEIG RONDA	20141	25	120	04	015	02980	1	0001	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)	P01	0001
PG	GRAN PASSEIG RONDA	20141	25	120	04	015	02980	1	0001	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)	P01	0003
PG	GRAN PASSEIG RONDA	20141	25	120	04	015	02980	1	0001	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)	P02	0001
PG	GRAN PASSEIG RONDA	20141	25	120	04	015	02980	1	0001	(null)	(null)	(null)	(null)	(null)	(null)	(null)	(null)	P02	0002

tipus_via	nom_via	pro_ine	mun_ine	codi_via_dgc	num	lletra	num_segons	lletra_segons	km	bloc	escala	planta	porta	parcela_catastral	num_ordre_bi
PS	RONDA GRAN	25	120	00119	0001		0000		00000		1	01	01	1394909CG0019C	117
PS	RONDA GRAN	25	120	00119	0001		0000		00000		1	01	02	1394909CG0019C	118
PS	RONDA GRAN	25	120	00119	0001		0000		00000		1	01	03	1394909CG0019C	119
PS	RONDA GRAN	25	120	00119	0001		0000		00000		1	02	01	1394909CG0019C	120

RET

Property Register

3. Validation. Horizontal Address (STREET)



FALSE POSITIVE

Only string similarity



RET			Property Register			Jaro Winkler similarity
Street code	Type of street	Name of street	Street code	Type of street	Name of street	
2512002980	PG	GRAN PASSEIG RONDA	2512000296	PS	SEGON DE RONDA	64
2512002980	PG	GRAN PASSEIG RONDA	2512000119	PS	RONDA GRAN	62
2512002980	PG	GRAN PASSEIG RONDA	2512001634	PS	DE LA TORRE PATRIOT	52
2512002980	PG	GRAN PASSEIG RONDA	2512000313	PS	RONDA	47
2512002980	PG	GRAN PASSEIG RONDA	2512001074	PS	ESTUDIANTS	45

TRUE POSITIVE

String similarity + spatial relations



RET			Property Register			Jaro Winkler similarity	% of intersection area
Street code	Type of street	Name of street	Street code	Type of street	Name of street		
2512002980	PG	GRAN PASSEIG RONDA	2512000119	PS	RONDA GRAN	62	96.69
2512002980	PG	GRAN PASSEIG RONDA	2512000313	PS	RONDA	47	0.74
2512002980	PG	GRAN PASSEIG RONDA	2512001634	PS	DE LA TORRE PATRIOT	52	0.00
2512002980	PG	GRAN PASSEIG RONDA	2512001074	PS	ESTUDIANTS	45	0.00
2512002980	PG	GRAN PASSEIG RONDA	2512000296	PS	SEGON DE RONDA	64	0.00

3. Validation. Vertical Address (FLOOR)

Sorting patterns have been defined to establish the equivalence RET - Property Register

Property Register	SA	Order 6				RET	
	AT, A1	Order 5		GROUP 4			
	+2	Order 4		GROUP 3	Order 4		A02, AT2, PSA, SAT
	+1	Order 3		GROUP 3	Order 3		A01, A1, PAT, ATC, A00, PA, ATI, A03
	n						Pn

	03, 3	Numerical Order		GROUP 2	Numerical Order		P03, P3
	02, 2						P02, P2
	01, 1						P01, P1
	PR, PPL, PP	Order 2			Order 2		PP, PRL, PPR, PPL, PAL
ES, EN, ENT	Order 1	GROUP 1	Order 1	ES, ENT, PEN			
0, BI, BJ, BE, B, B0...	Order 0		Order 0	PBE, PBI, BI, PBJ, PO, PBA, PLB, B1...			
SM, SS	Order -1		Order -1	SSO, SS, SST			
S1, -1, SO, ST	Order -2	GROUP -1	Order -2	SOT, S01, PSO, S1, PS1			
S2, -2	Order -3		Order -3	S02, PS2, S2			

3. Validation. Vertical Address (STAIRCASE and DOOR)

The address is broken down into **numbers** and **words** and a search is made for equivalences regardless of the order of the parts.

For fields not in the table of correspondences, a **character similarity algorithm** is used.

Examples of correspondences

STAIRCASE		DOOR	
RET	Property Register	RET	Property Register
01	1	0001	01
01	A	0001	A
01	01	001B	B1
A	1	B	02
02	B		
D	4		

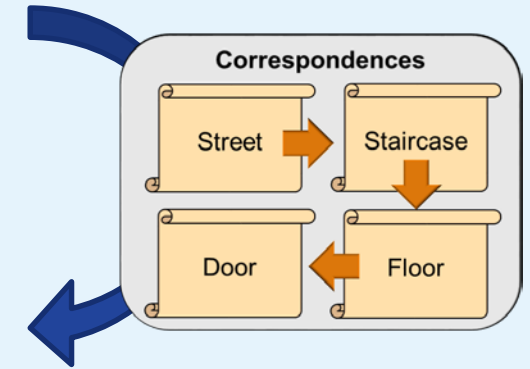
3. Validation. Full address

RET

Province	Municipality	Street code	Type of street	Name of street	Street number	Staircase	Floor	Door
25	120	02980	PG	GRAN PASSEIG RONDA	0124	B	P07	C

Property Register

Province	Municipality	Street code	Type of street	Name of street	Street number	Staircase	Floor	Door
25	120	00119	PS	RONDA GRAN	0124	2	07	03



Results (% found)

	Horizontal Address	Vertical Address
Barcelona	96.7%	89.8%
Catalonia	89.7%	76.0%

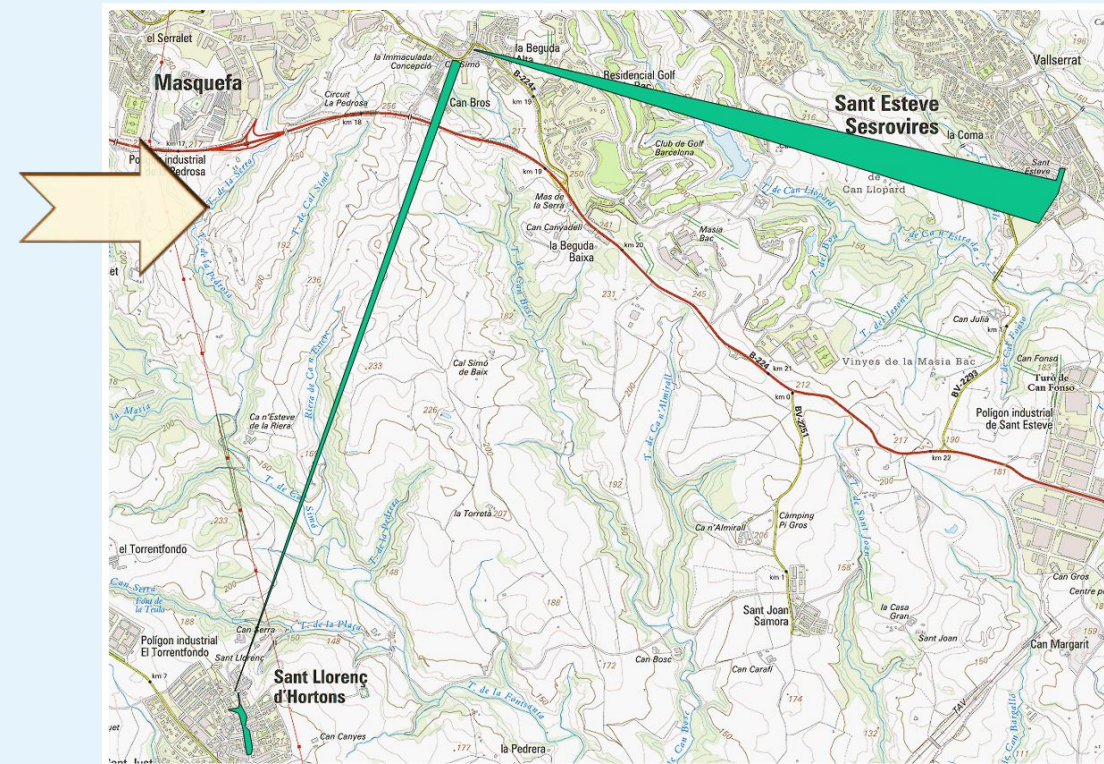
4. Quality control. Horizontal Address

Problem:

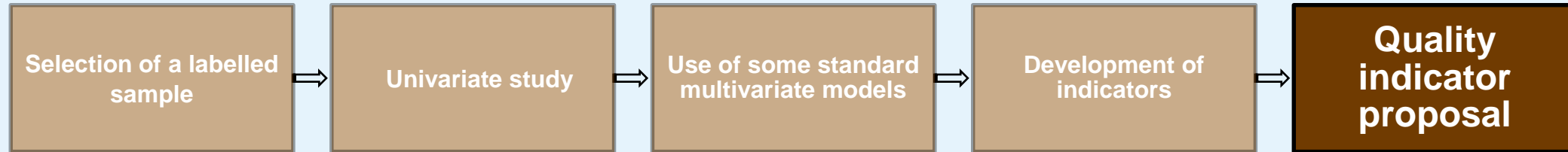
The construction of polygons using all the points on a street (street addresses) produces anomalous points (erroneous polygons).

Analysis:

In order to **detect the maximum number of erroneous polygons**, it is necessary to find an instrument to quantify the quality of this spatial data. The process used is as follows:



4. Quality control. Horizontal address



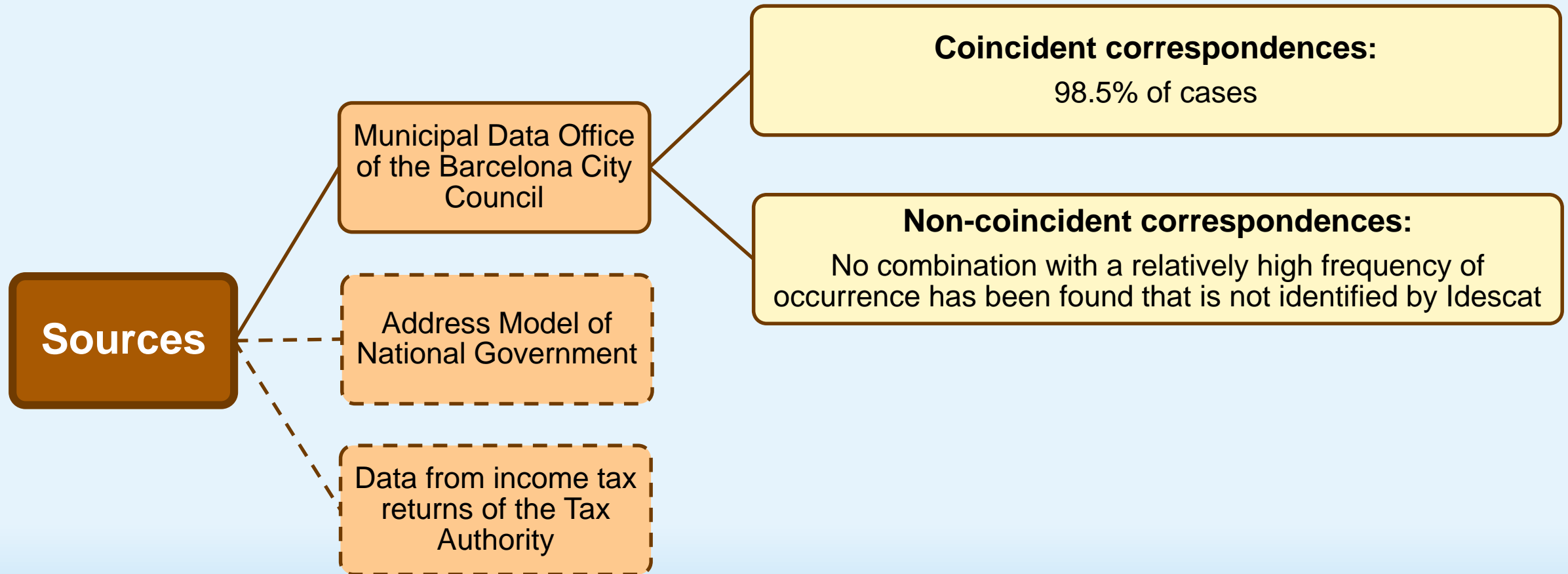
$$PQI_p = CV_p^2 \cdot \max\{length_i - length_{i+1}\}_{i=1 \dots v_p - 1}$$

Where “i” iterates along the sides of the polygon according to its decreasing length and “CVp” is the variation coefficient of the side’s lengths.

This indicator allows us to **classify** and **rank** polygons by **estimated quality** and so identify possibly erroneous polygons.

4. Quality control. Vertical Address

Comparison of Idescat's tables of equivalences with different sources:



5. Conclusions

- The Statistical Territory Register of Catalonia incorporates different sources for the purpose of assigning and validating a **territorial identifier** for each of the **statistical units**.
- The **diversity** of the descriptions of the postal addresses in the sources hinders the proper spatial identification of the units.
- The use of spatial and geometrical methods has been a very powerful tool for establishing **tables of equivalences** between the sources, as well as **validating, detecting errors** and **establishing quality indicators** for the georeferenced addresses.