



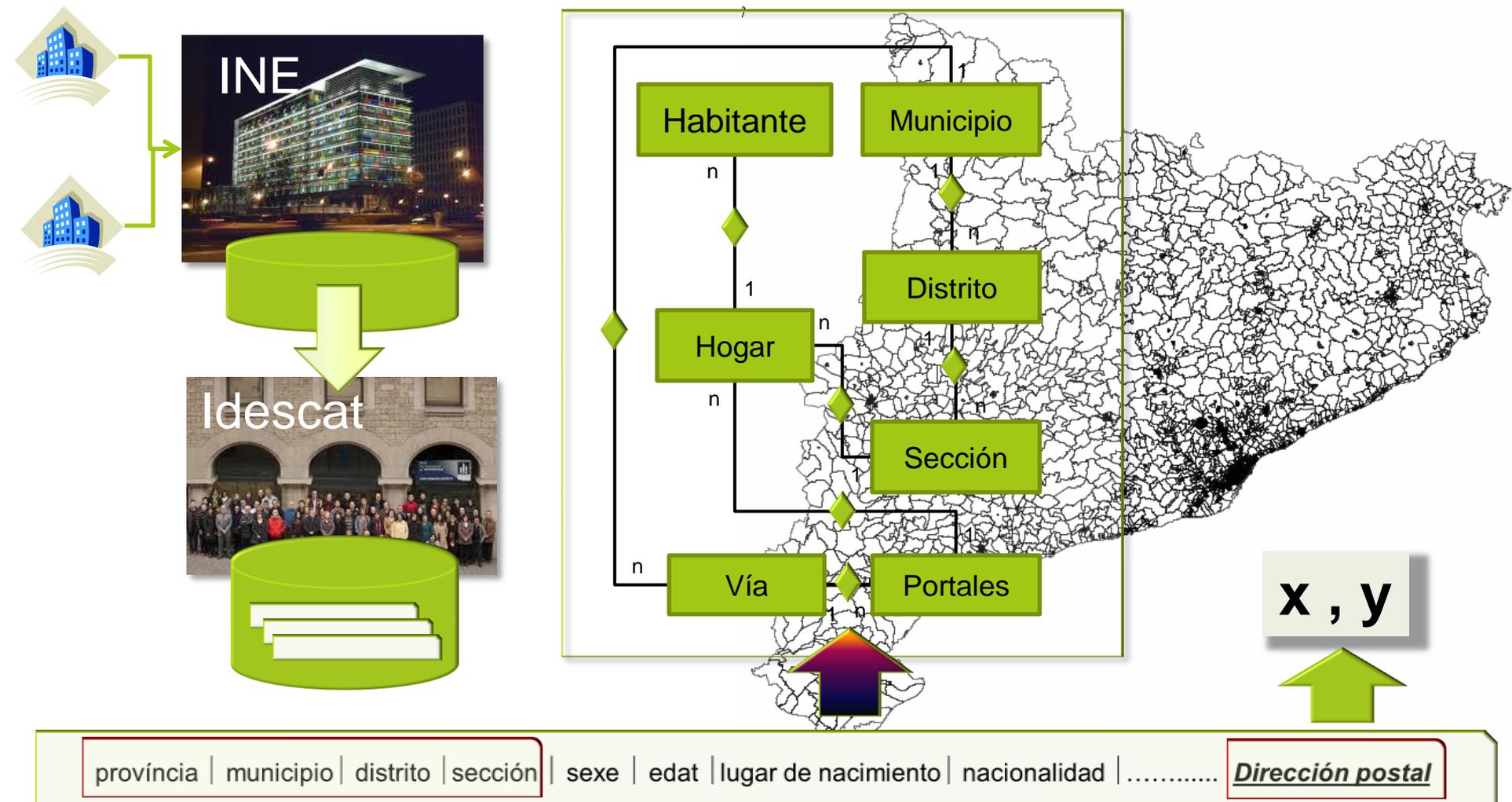
Generalitat de Catalunya
Institut d'Estadística de Catalunya

Hacia el Registro estadístico de territorio

- La geocodificación del Registro de población 2012, 2013 i 2014
- Imputación de coordenadas
- Hacia un sistema de producción basado en registros administrativos
- Registre estadístico de territorio. Finalidades.
- Fuentes : Catastro, Cartociudad, Censo de Edificios 2011, AGE, BDMAC
- Solución...

La geocodificación del Registro de población 2012, 2013 i 2014.

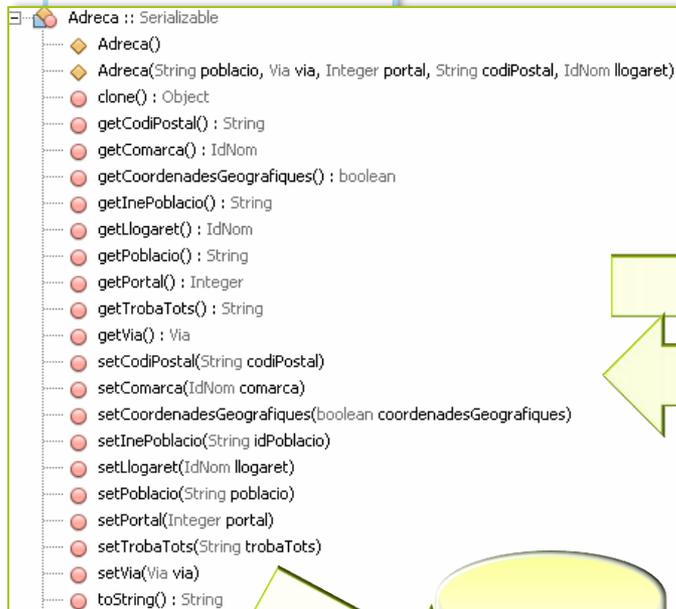
Información de partida. Registro de población.



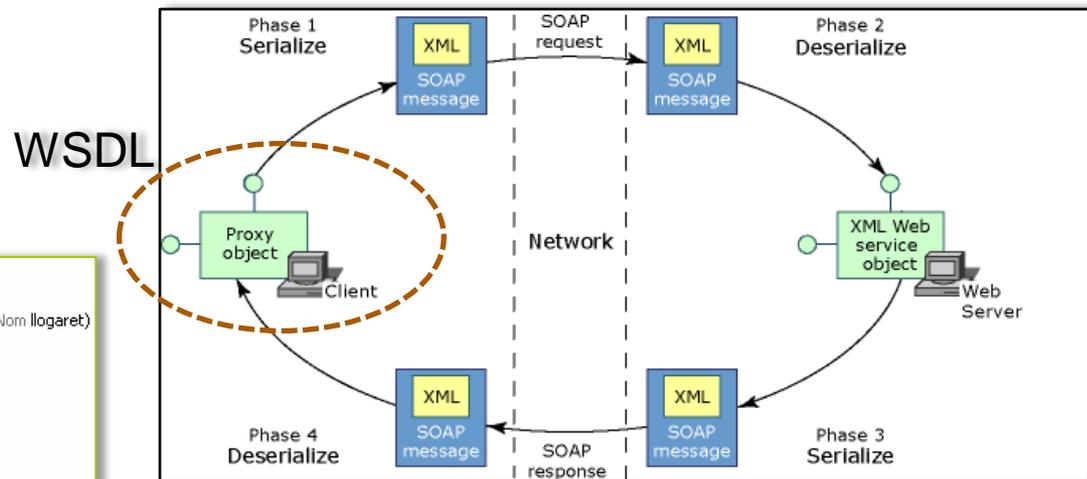
Geocodificador ICGC. Resultados. RP 2012, 2013, 2014.

Retreiver.java
RetreiverAdreca.java

Punt.java
Via.java
Adreca.java



BD
espacial



(x,y), epsg: 25831, indicador exactitud...

Any	Portals	Geocodificados	% geocodificados	% no geocodificados
2012	992.660	918.192	92,50	7,50
2013	990.000	914.316	92,36	7,64
2014	1.038.595	968.166	93,22	6,78

Qualitat. Geocodificació RP2014.

Qualitat	Barcelona	Girona	Lleida	Tarragona	Catalunya
Portal	207942	37086	28785	49630	323443
Portal interpolado	284200	90343	40047	74140	488730
Portal interpolado más cercano	89931	19617	8889	19364	137801
Rectángulo relativo a la calle	8128	4920	2420	2724	18192
No resuelto	34098	12602	11672	12057	70429
% no resuelto	5,4	7,7	12,7	7,6	6,8
Total	624299	164568	91813	157915	1038595

- El servicio mejora en la su calidad año tras año

Exactitud	2013	2014	2014 Barcelona capital
Portal	-	31 %	70 %
Portal interpolado	~ 70 %	47 %	28 %

Imputación.

- Para los portales no geocodificados es necesario imputar la su posición
- No sirve cualquier posición:
 - Los habitantes tienen a priori una sección censal asignada
 - Tiene que ser razonables (nadie vive en un campo de patatas o es muy poco probable)
- La imputación, finalmente, dependerá de los datos disponibles

2012 i 2013

- Geometría secciones censales
- SIGPAC
- Fotografías satélite NOAA

2014

- Catastro
- SIGPAC

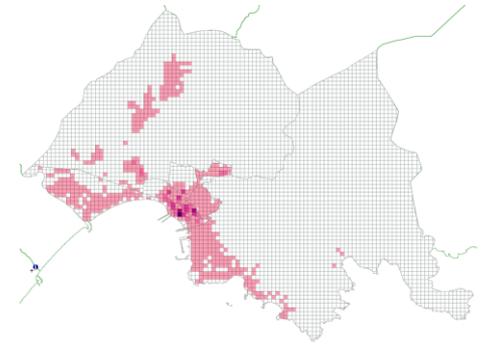
~~Geometría de las secciones censales~~

Imputación 2012 i 2013.

$[x,y]=ST_PointOnSurface(R[k])$

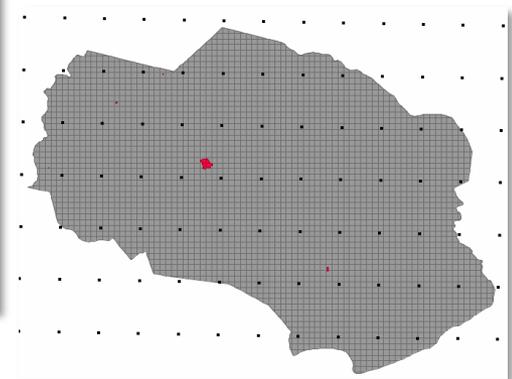
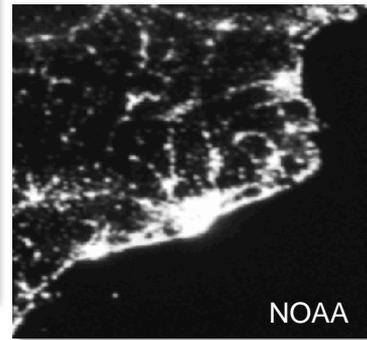
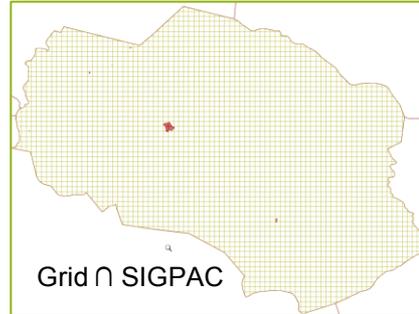


Si % geocod. > 70 % $p[k] \sim f(R[k])$

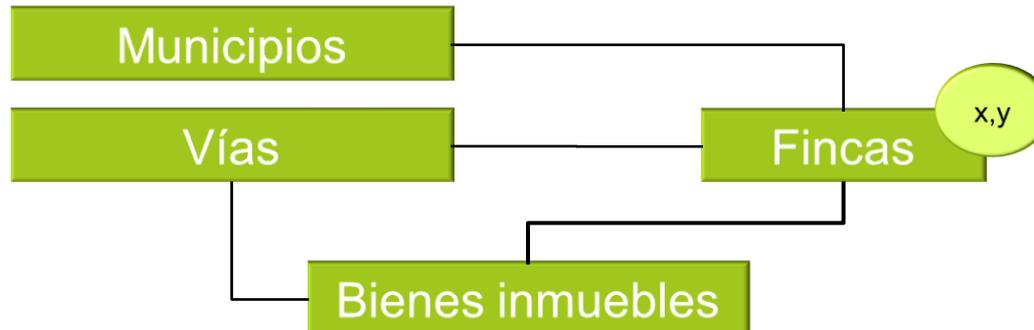


En caso contrario $p[k] \sim L[k]*s[k]/\sum L[k]*s[k]$

¿Cómo estimar $p[k]$?



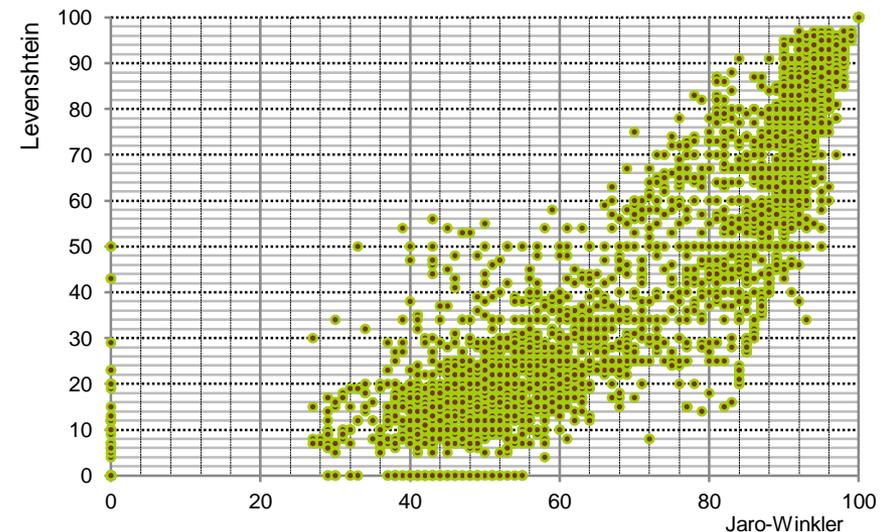
- Disponemos de la información relativa al archivo 'CAT' del Catastro



- No disponemos del seccionado (geometría)
 - estrategia de asignación/imputación diferente
- Disponemos de las posiciones de los centroides de las fincas
- y de los códigos de vía, nombres de vías, de las numeraciones...
- Pero : Catastro y INE utilizan un conjunto de códigos diferentes

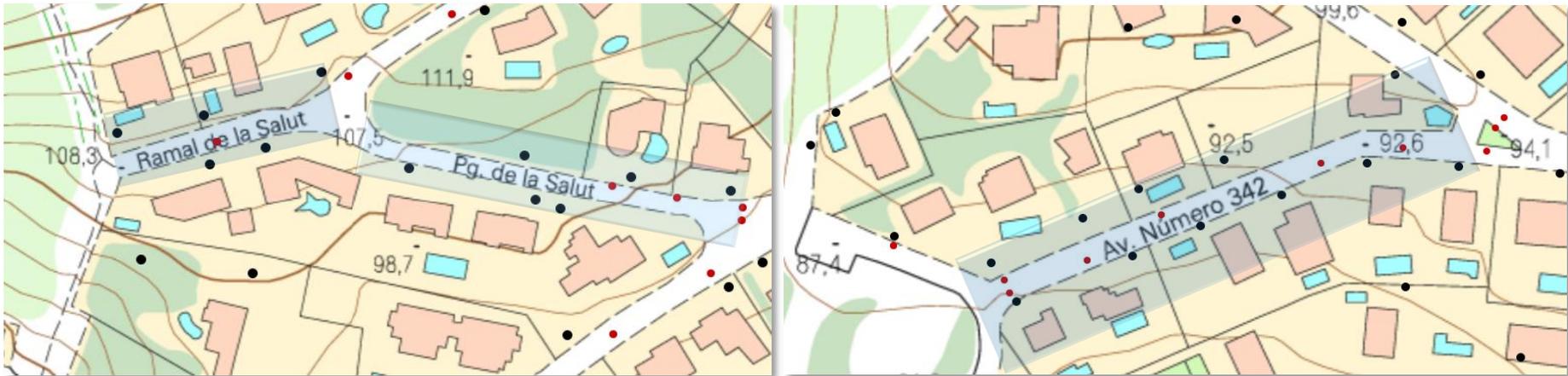
Imputación. RP2014. Fuzzy search.

- Aproximadamente sólo un 50% de les vías INE i DGC tienen el mismo nombre
- Por tanto es necesario
 - Utilizar una métrica
 - Escoger un límite por encima del cual supondremos que vía INE = vía DGC
- Métricas disponibles
 - Levenshtein
 - Jaro-Winkler
 - Trigrams
 - etc..
- Jaro-Winkler i s=90



Imputación. RP 2014. Relación entre posiciones portal geocodificado y finca catastro.

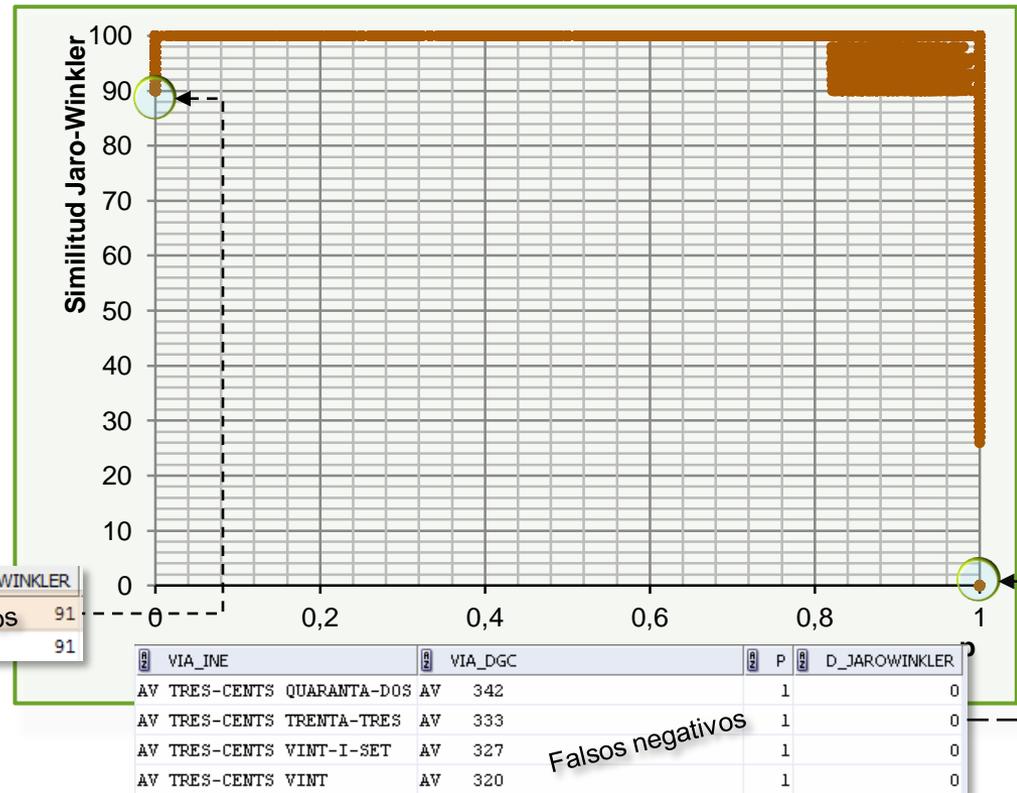
- A cada portal geocodificado, p1, asignemos la finca **más cercana** p2
 - Cada par (p1,p2) tiene implícito un par (vía INE, vía DGC)
 - Calculamos la frecuencia relativa, f, de los pares (vía INE, vía DGC) para todos los puntos y si...
 - Si $f \geq U$ para un par (vía INE, vía DGC) \rightarrow vía INE = vía DGC



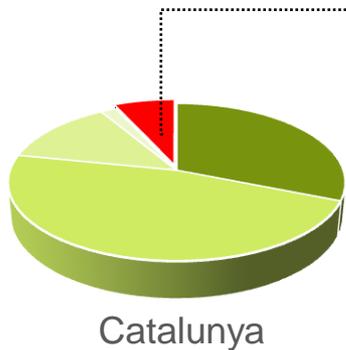
Vía INE	Vía DGC	f	Jaro-Winkler
AV TRES-CENTS QUARANTA-DOS	AV 342	1	0
RAMAL SALUT	C RAMAL DE LA SALUD	1	50
PG SALUT	PS SALUT, DE LA	1	88
			Falsos negativos con fuzzy search

Imputación. RP2014. Criterios de asignación.

- ✓ Similitud Jaro-Winkler ≥ 90
- ✓ $f \geq 0.85$
- Cubre un 76 % de vías INE



TVIA	NVIA	TIPUS_VIA	NOM_VIA	D_JAROWINKLER
D C	CAN MASET	DS	CAN MARTI	91
D CAMI	ANTIC DE GRANOLLERS	CM	ANTIC GRANOLLERS	91



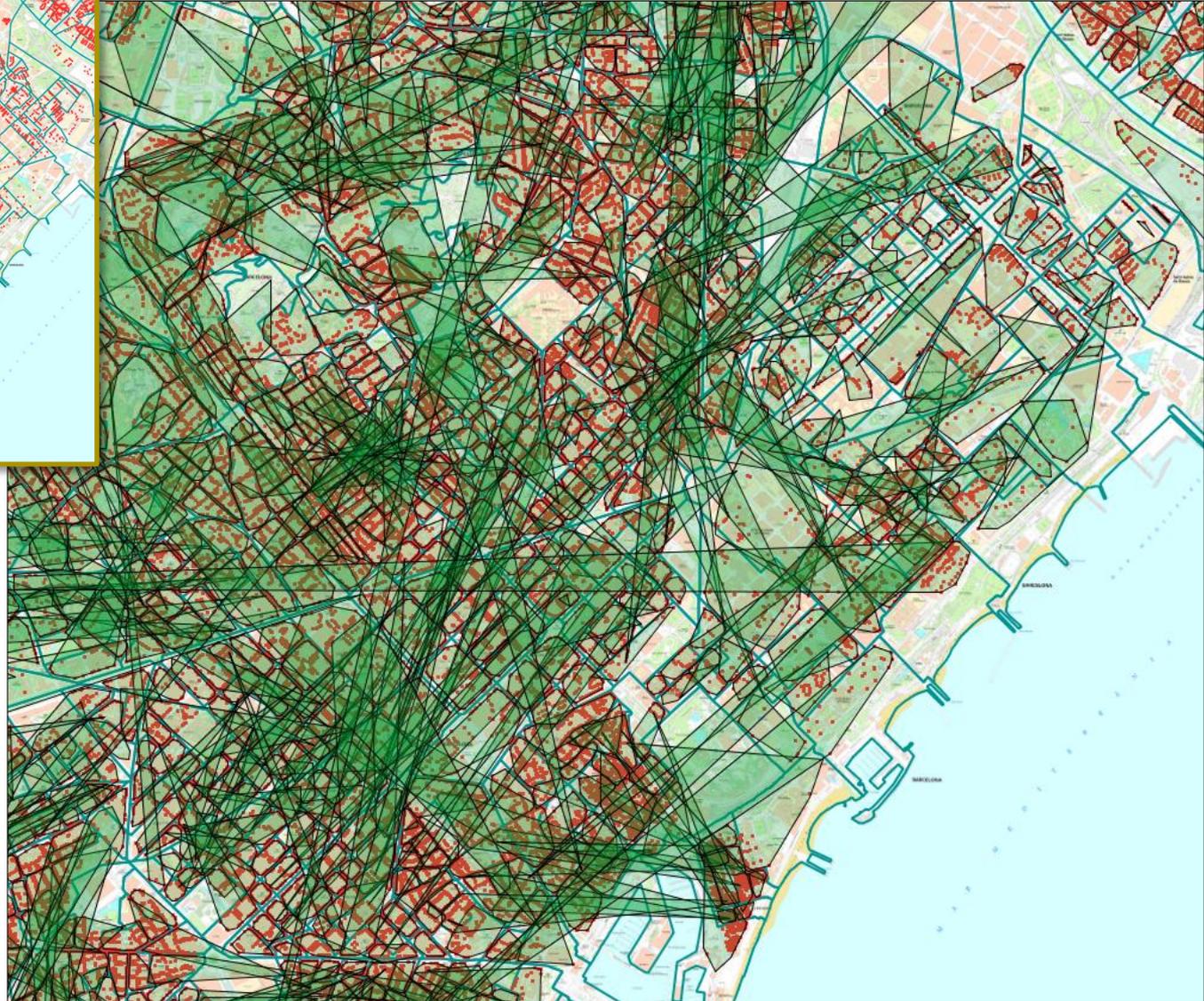
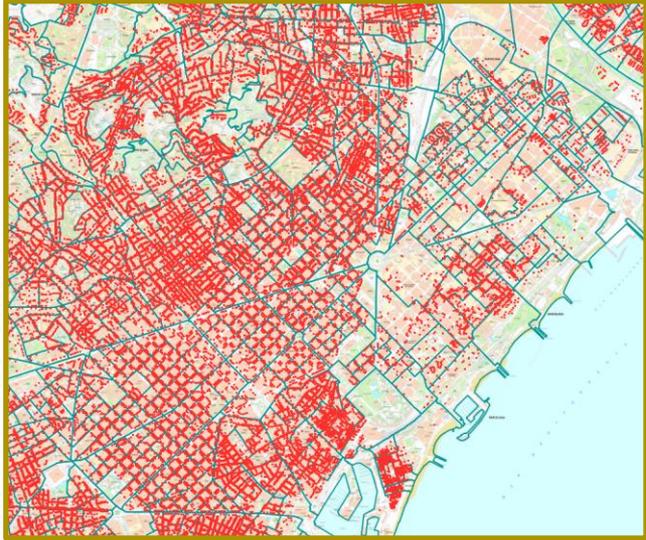
A. \exists la vía en la taula i coincide la numeración

B. \exists la vía en la taula i no coincide la numeración

C. No \exists la vía en la taula

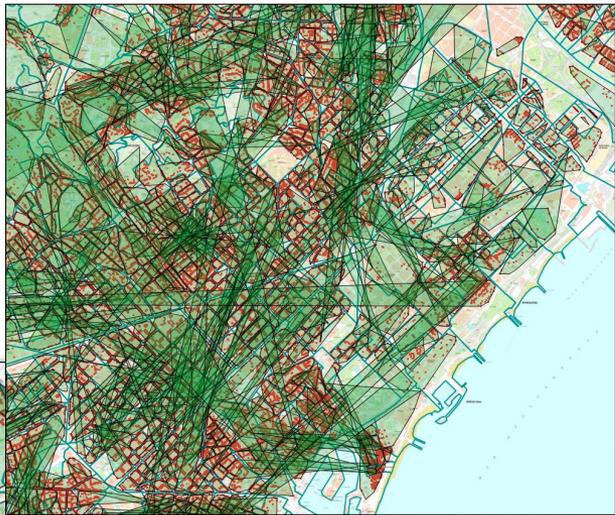
Delimitar el espacio
 No disponemos de la cartografía de
 las secciones censales para 2014

Imputación. RP2014. Aproximación a les secciones censales mediante un *convex-hull*

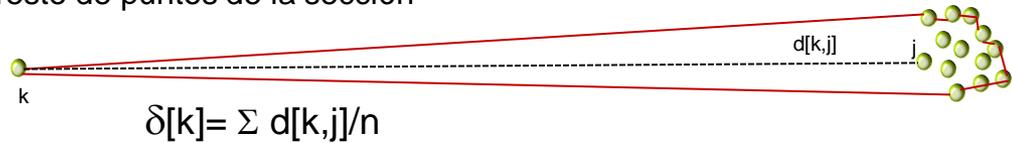


y el resultado es.....

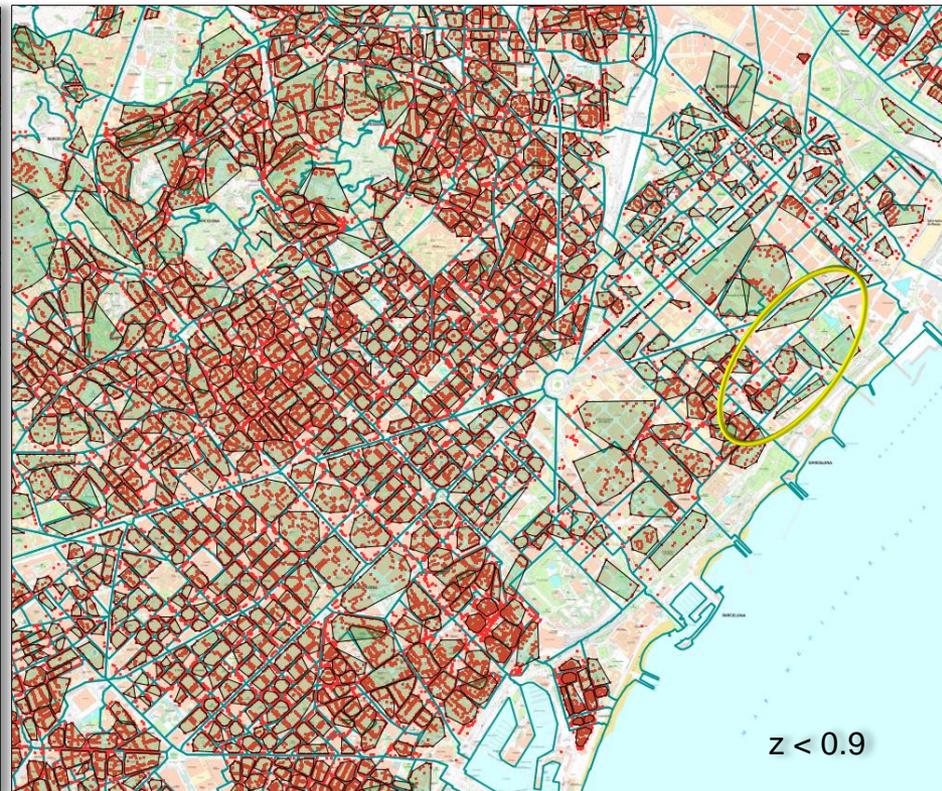
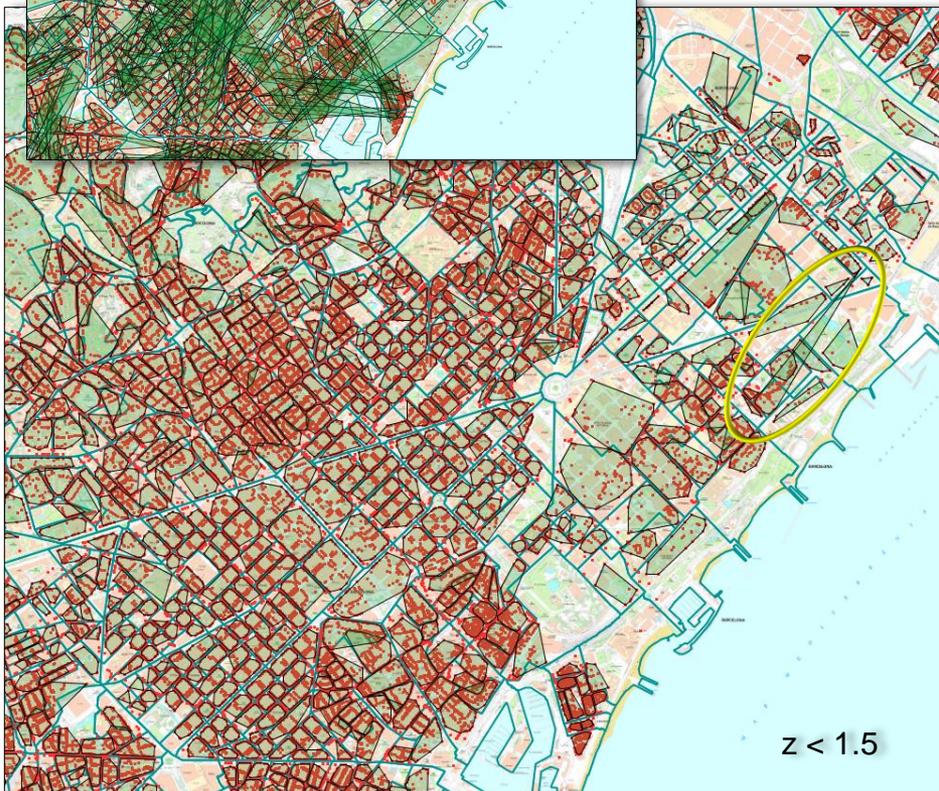
Imputación. RP2014. Aproximación a las secciones censales.



Algunos puntos son *outliers* en la distribución de las medias de las distancias respecto del resto de puntos de la sección



Obteniendo el z-score de estos valores, rechazamos aquellos puntos con un valor que está per encima de un cierto límite en la construcción del convex.



RP2014. Zonas de imputación.

- Para municipios con más de una sección censal

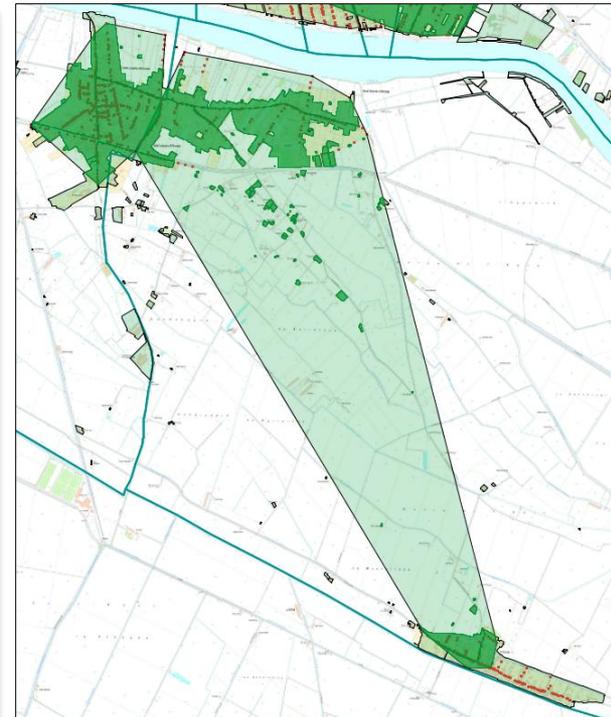
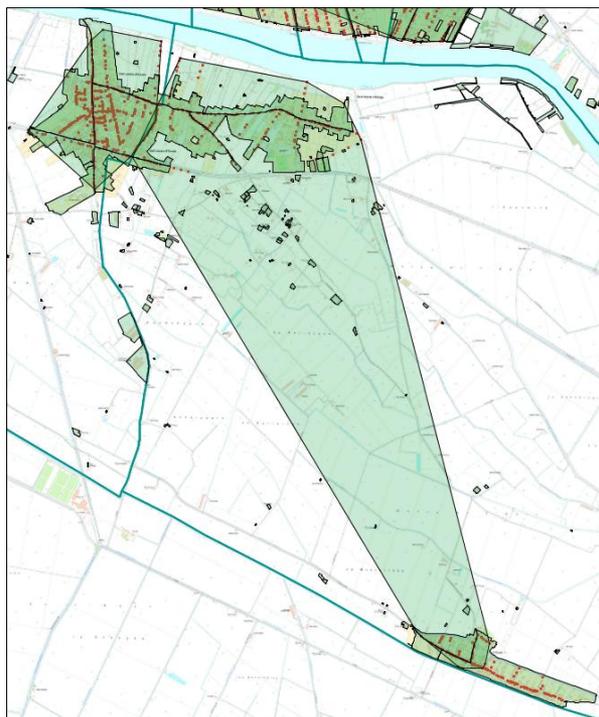
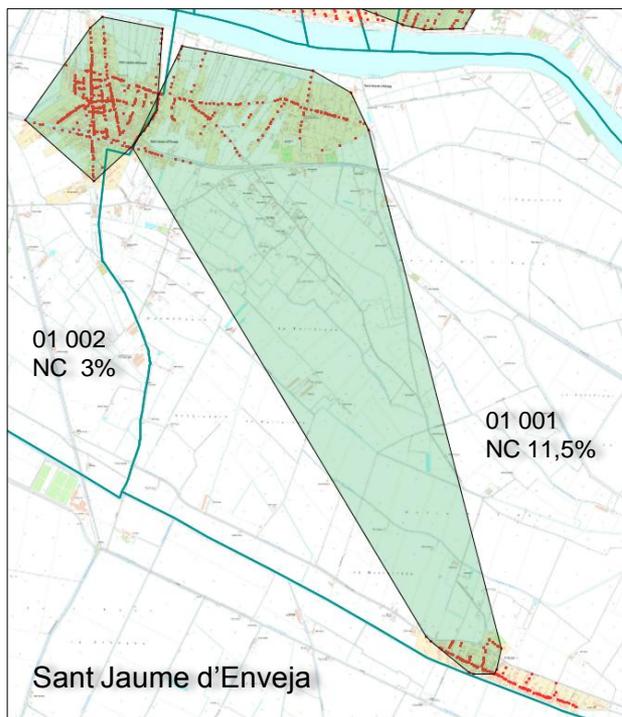
convex hull $z < 0.9$



SIGPAC zona urbana

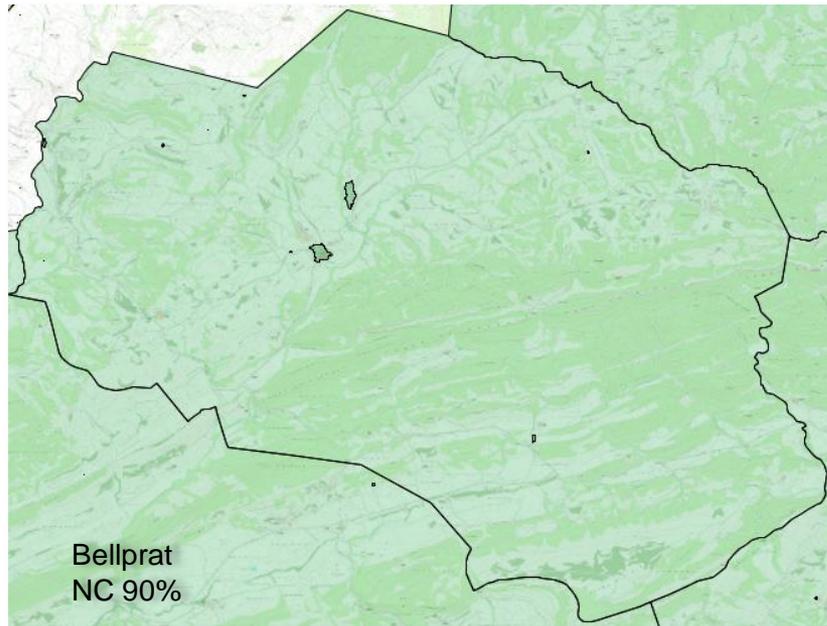


zona de imputación



RP2014. Zonas de imputación.

- Para municipios con una sección censal



zona de imputación = límites del municipio \cap SIGPAC zona urbana

RP2014. Esquema de imputación.

portal no geocodificado NC (6,78%)
74.415 portales

existe la vía en la tabla de correspondencias INE-DGC

existe una finca con numeración idéntica

0

asignación directa

no existe una finca con numeración idéntica

-1

finca con numeración cercana dentro de zona de imputación

-2

finca con numeración cercana sin restricción espacial

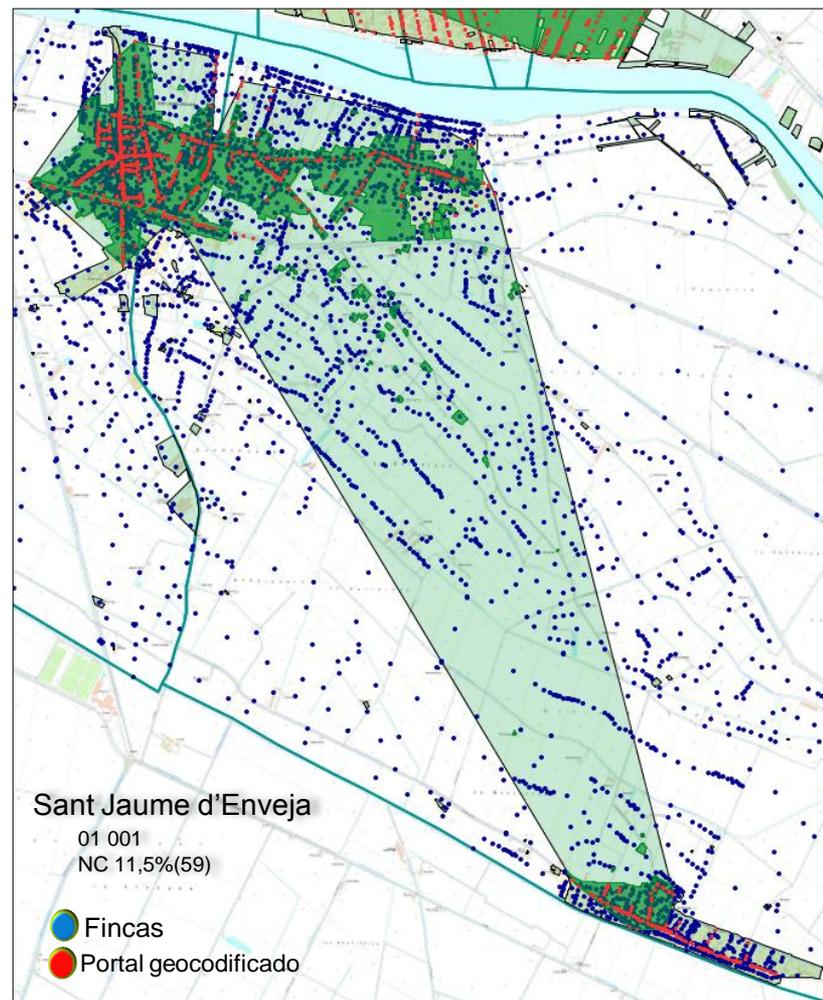
no existe la vía en la tabla de correspondencias INE-DGC

-3

aleatoriamente a una finca dentro de zona de imputación

-4

¿p[k] ?



RP2014. Esquema de imputación.



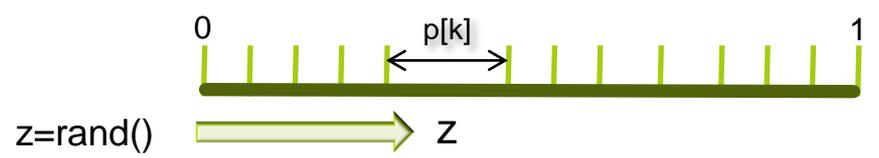
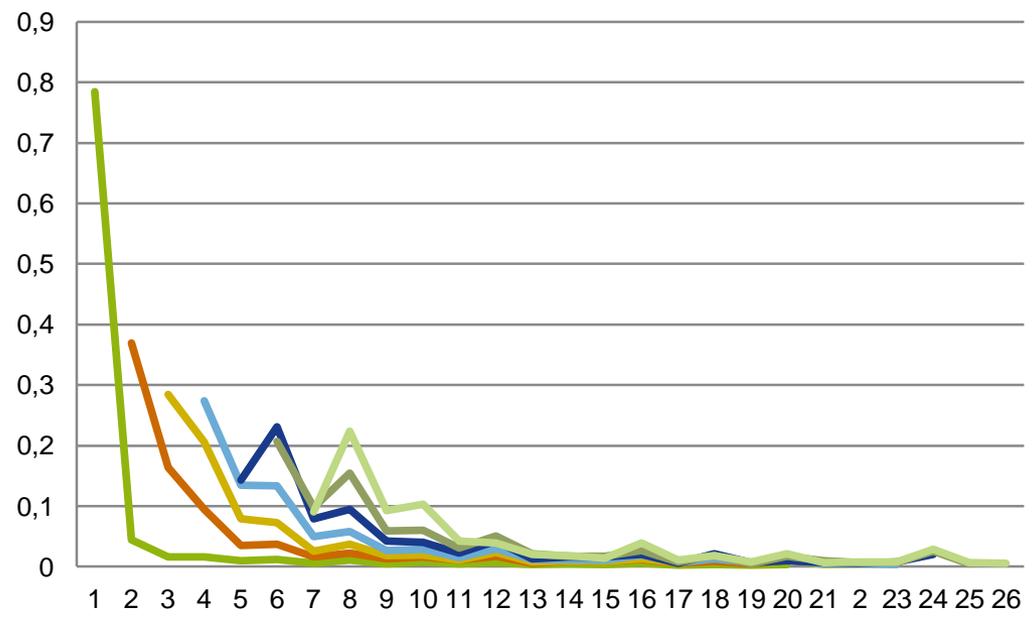
- Fincas más cercanas
- Portal geocodificado
- Fincas donantes

no existe la vía en la tabla de correspondencias INE-DGC

-3

aleatoriamente a una finca dentro de zona de imputación

$p[k]$ ~ frecuencia relativa nº hogares frente nº bien inmueble pares portales - fincas más cercanas



RP2014. Imputación. Resultados.

portal no
geocodificado

NC (6,78%)
74.415 portales

Existe la vía en la tabla de correspondencias INE-DGC

Existe una finca con numeración idéntica

0

asignación directa

no existe una finca con numeración idéntica

-1

finca con numeración cercana dentro zona de imputación

-2

finca con numeración cercana sin restricción espacial

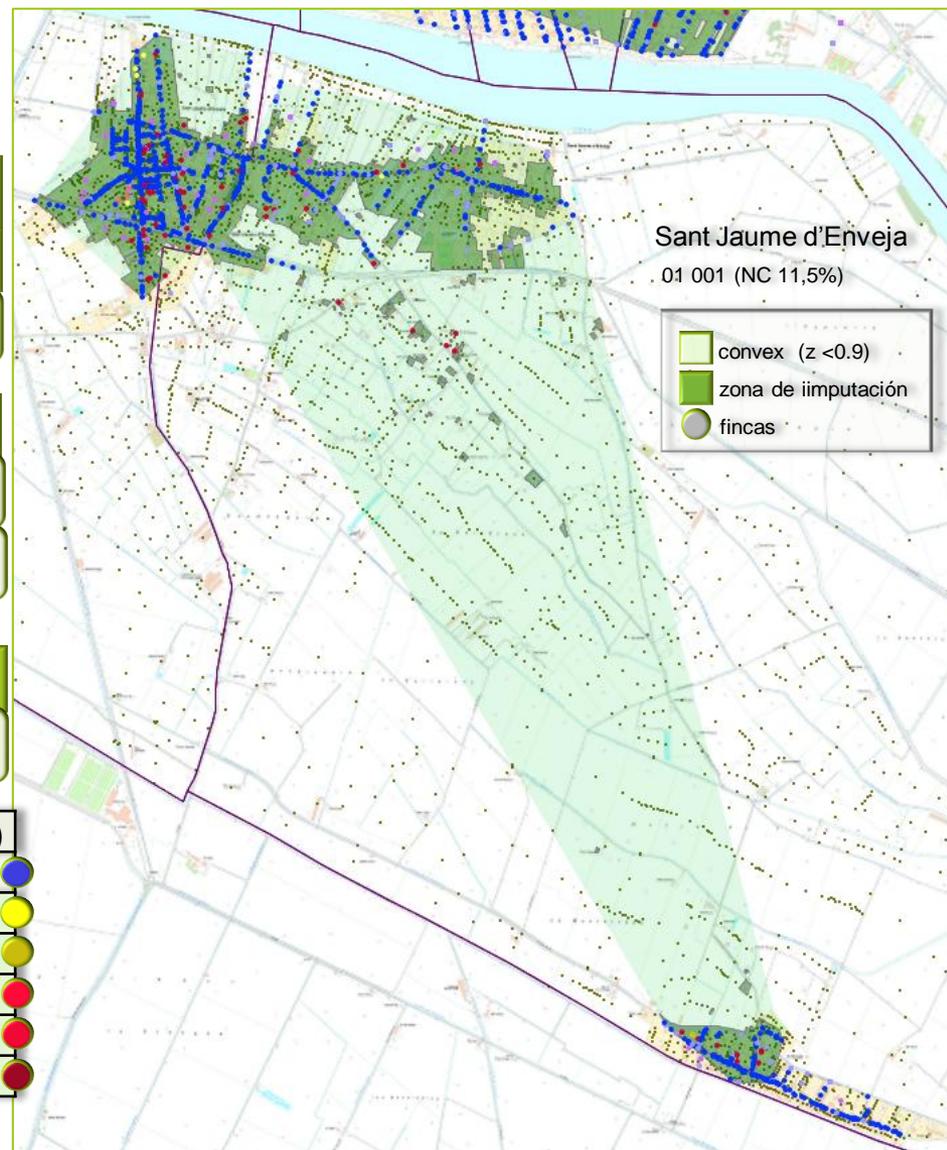
no existe la vía en la tabla de correspondencias INE-DGC

-3

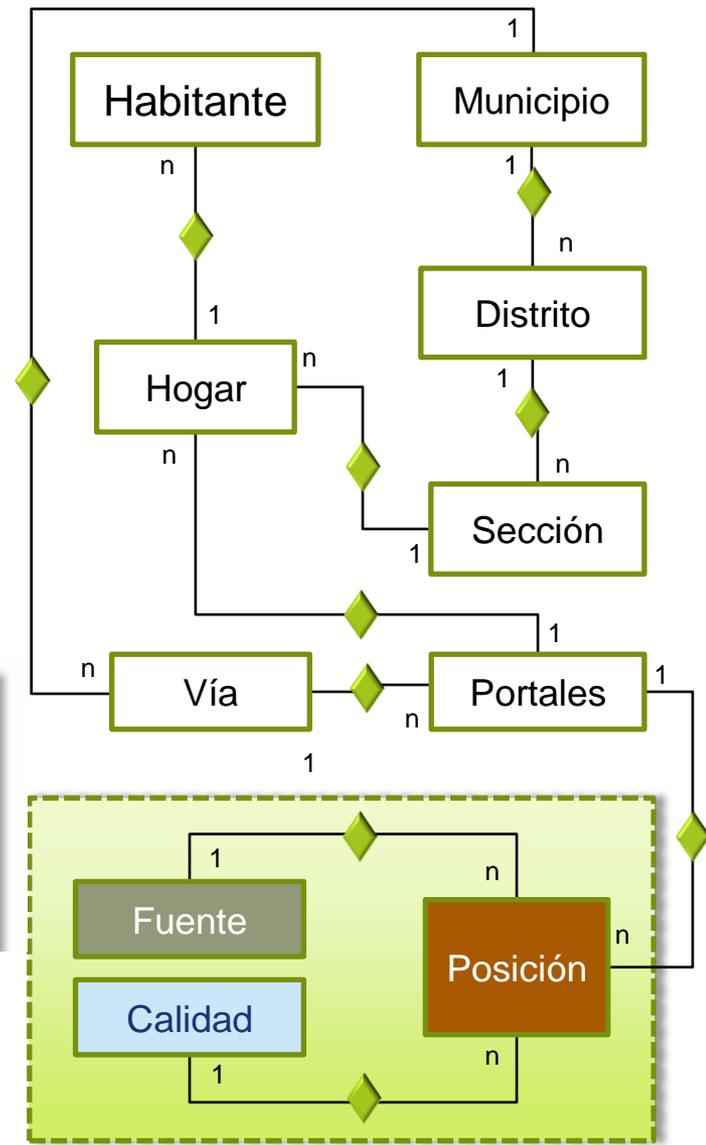
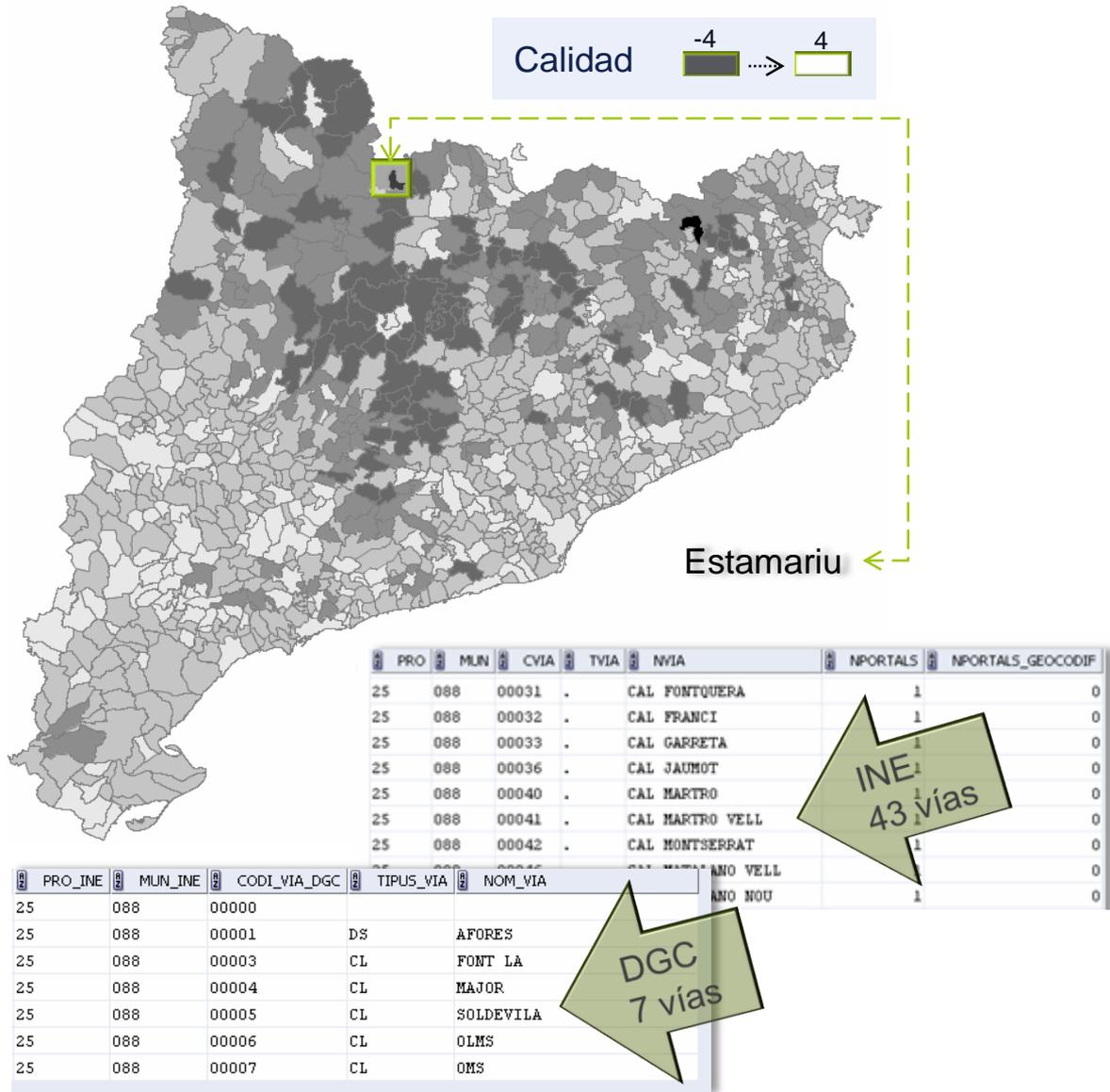
-4

aleatoriamente a una finca dentro de zona de imputación

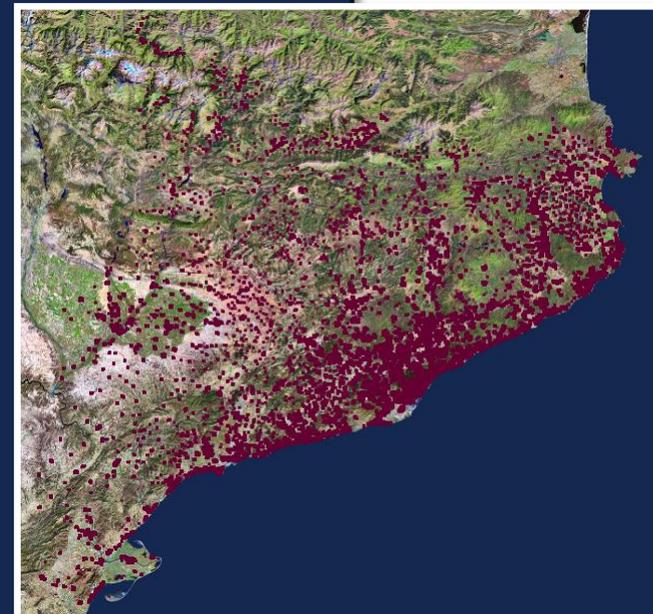
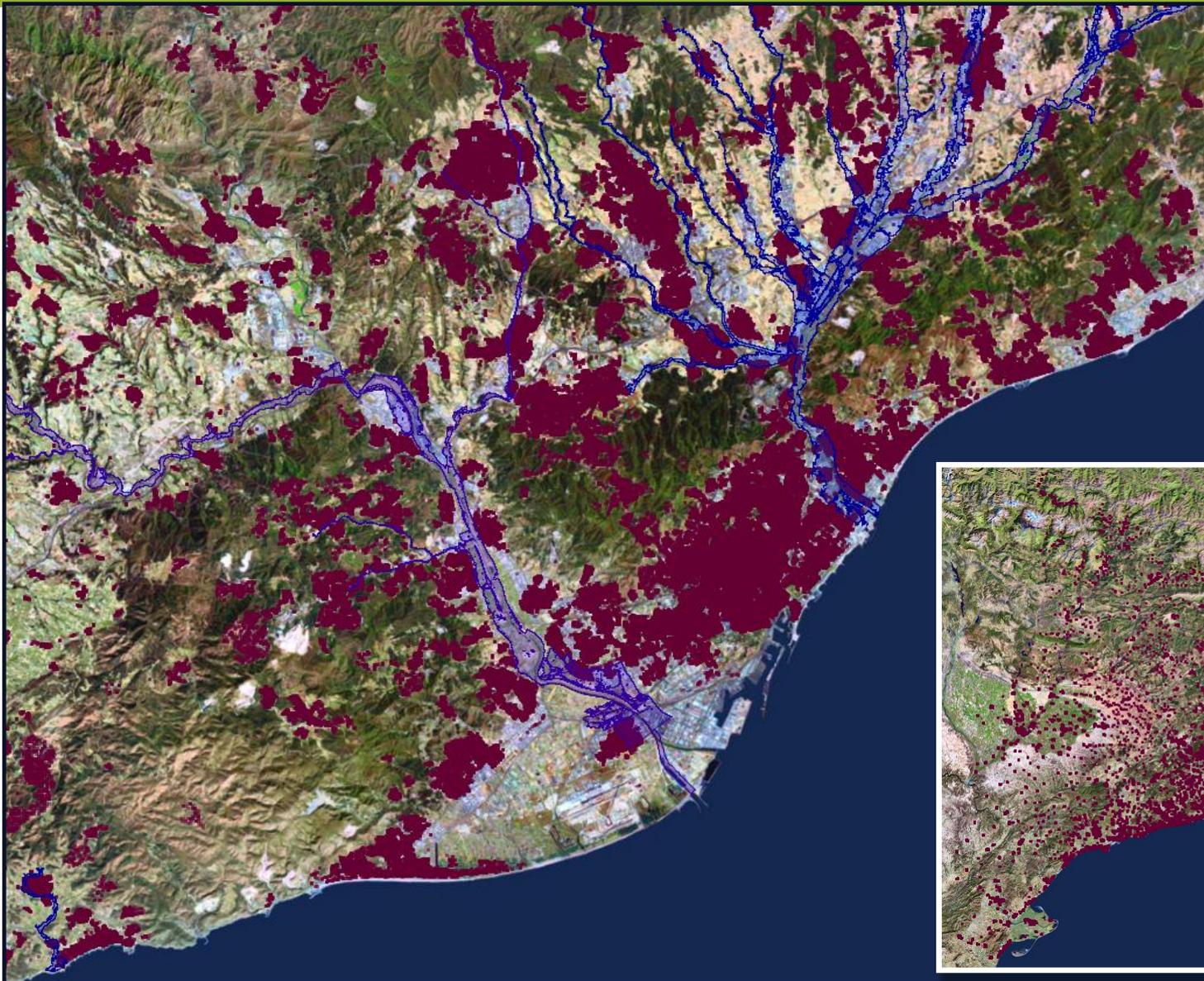
Tratamiento	% (portales)	%(habitantes)
Geocodificados	93.22	95,58
0	2.14	1,68
-1	0.22	0,11
-2	1.13	0,91
-3	2.98	1,50
-4	0.31	0,21



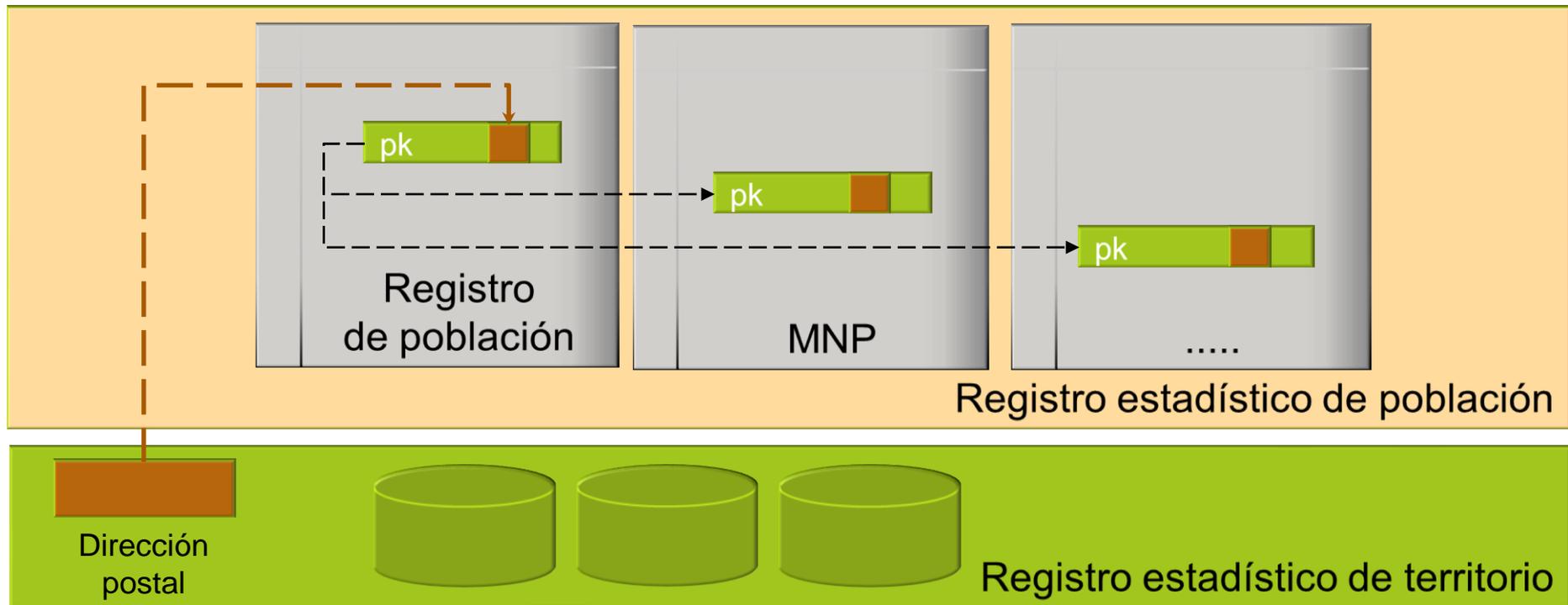
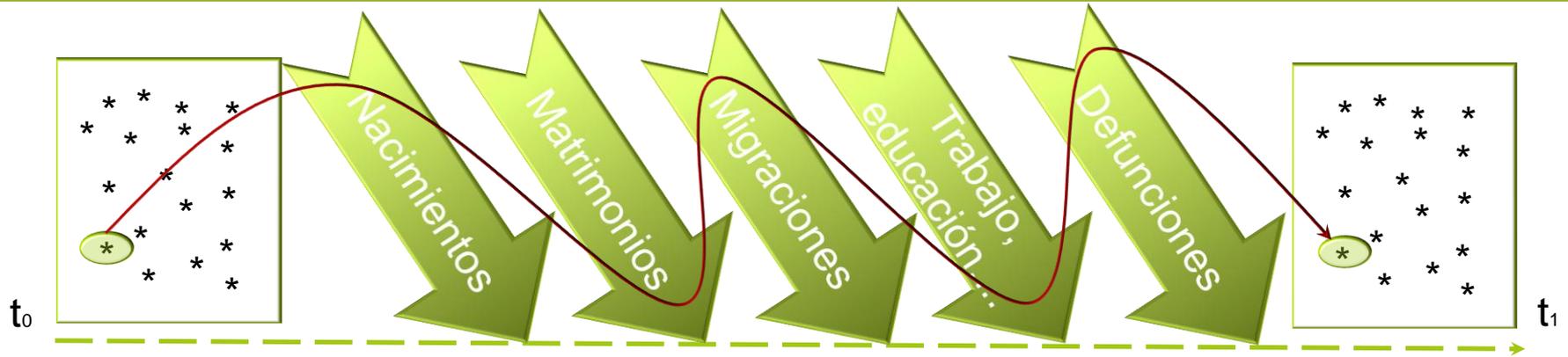
RP2014. Imputación. Resultados.



RP2014. Resultados.



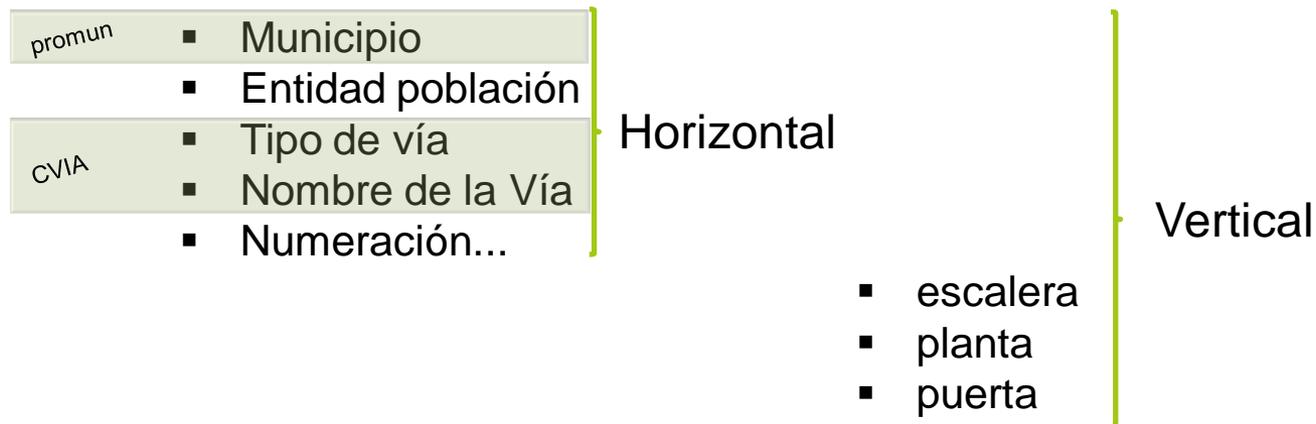
Hacia a un sistema de producción basado en registros administrativos.



Registro estadístico de territorio.



- Una dirección se construye generalmente con un conjunto de campos



Carrer Diputació 151, 3r, Barcelona (→ 080193 01132 151 3r)

- No siempre es así:
 - *N340 punto kilométrico 214*
 - *Cal Masianes, Organyà*

Registro estadístico de territorio. Finalidades.

❖ Validar una dirección

¿Existe en el conjunto finito de direcciones de Catalunya?

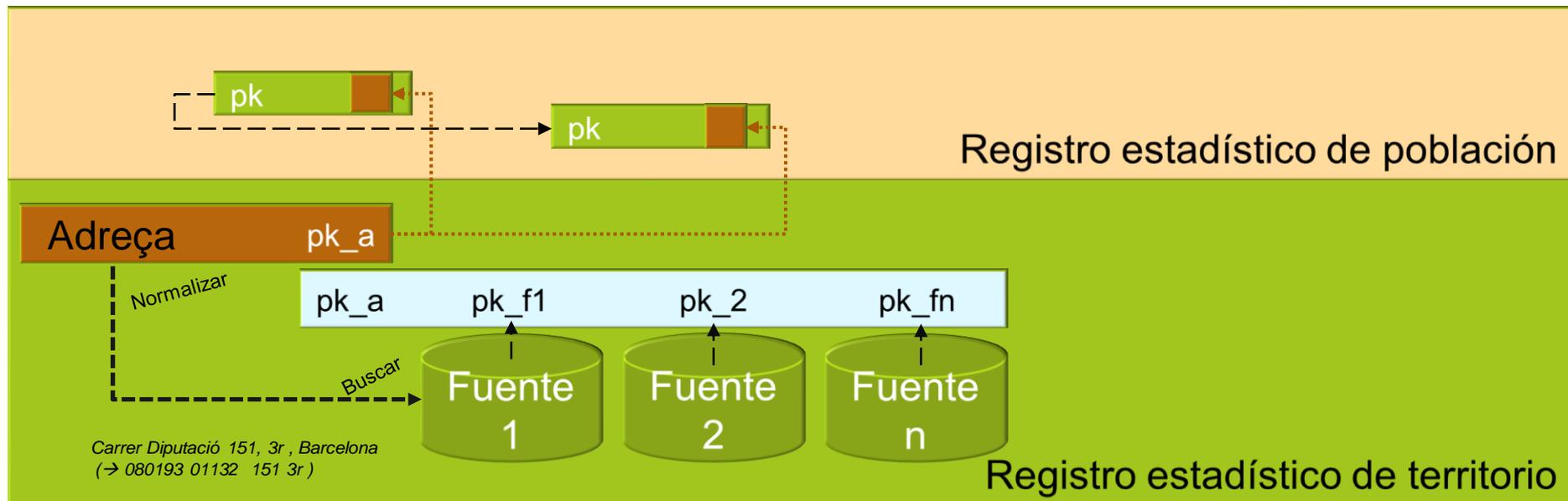
Similar a comprobar que un código de ocupación existe en la clasificación oficial

Pero, existen diferentes fuentes....

❖ Asignar una pk (address code) a cada dirección i propagarla al REP

Crear una pk es siempre factible: un número secuencial es suficiente.

Pero tiene más sentido si la dirección es válida i per tanto se puede relacionar con la pk de la fuente contra la que validamos.

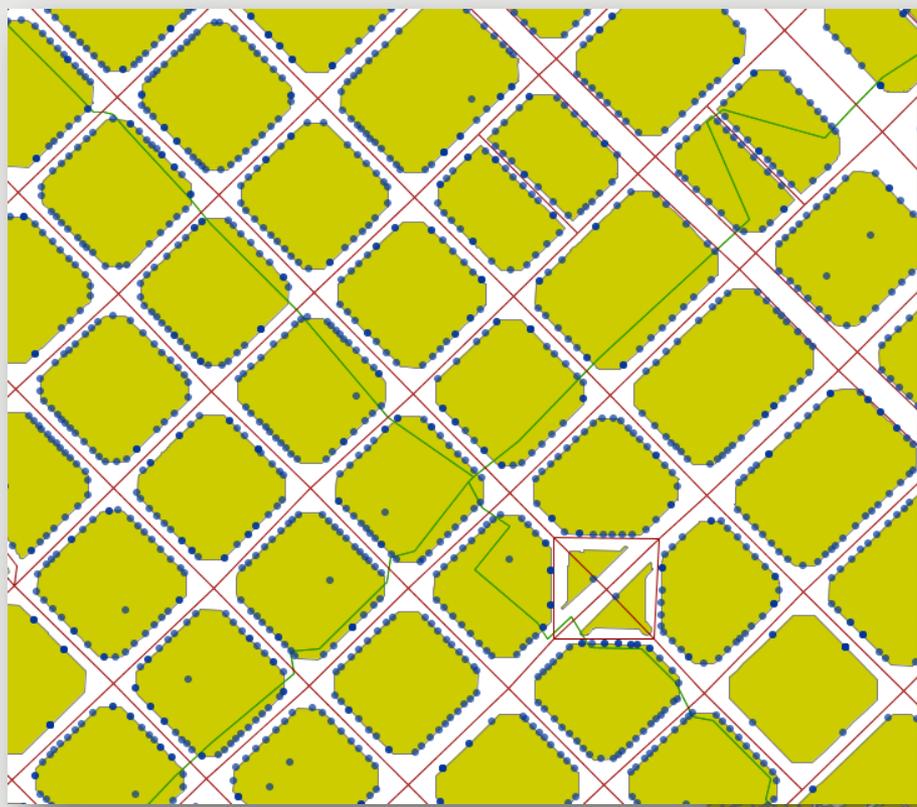


RET. Fuentes.

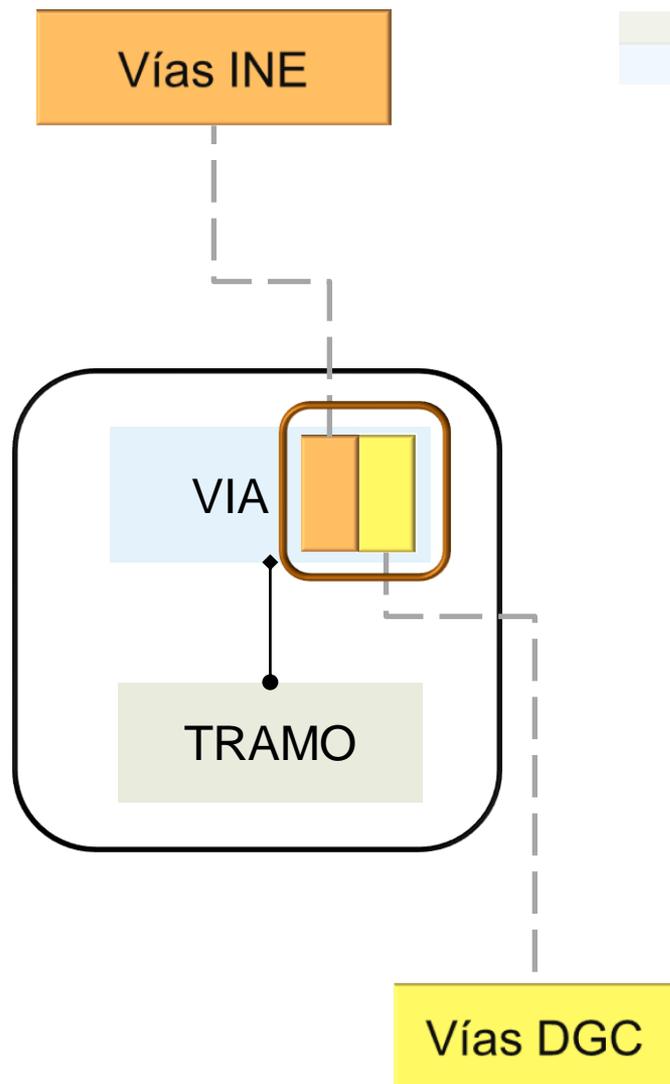
- ✓ Exhaustiva
- ✓ Calidad

Fuente	Organismos	Observaciones
Catastro. Archivo CAT	DGC	Modeliza fincas, bienes inmuebles, unidades constructivas etc. <u>Códigos propios, diferentes a INE</u> Fuzzy search.... falsos positivos!!
Cartociudad	IGN (INE;DGC;Correos)	Modeliza vías, tramos de vías, portales, códigos postales
Censo Edificios 2011. Aproximaciones postales.	INE	Contiene información sobre las aproximaciones postales de los edificios incluidos en la <u>operación Censo 2011</u>
Direcciones AGE	INE, IGN, DGC, Correos	Modeliza direcciones. <u>En fase de diseño...</u>
BDMAC	ICGC, municipios, IDESCAT	Modeliza direcciones. Modelo aprobado C4

RET. Fuentes. Cartociudad. Contenido.



RET. Fuentes. Cartociudad. Tramo-Vía



	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	ID_TRAMO	NUMBER	Yes	(null)	1	(null)
2	ID_VIAL	NUMBER	Yes	(null)	2	(null)
3	INE_VIA	VARCHAR2(10 BYTE)	Yes	(null)	3	(null)
4	DGC_VIA	NUMBER(10,0)	Yes	(null)	4	(null)
5	TIPO_VIA	NUMBER(5,0)	Yes	(null)	5	(null)
6	TIPO_V_DES	VARCHAR2(100 BYTE)	Yes	(null)	6	(null)
7	TIP_VIA_IN	VARCHAR2(25 BYTE)	Yes	(null)	7	(null)
8	NOM_VIA	VARCHAR2(100 BYTE)	Yes	(null)	8	(null)
9	NOM_ALTERN	VARCHAR2(100 BYTE)	Yes	(null)	9	(null)
10	NOM_AL_DES	DATE	Yes	(null)	10	(null)
11	V_FUENTE	VARCHAR2(100 BYTE)	Yes	(null)	11	(null)
12	V_FUENTE_D	VARCHAR2(100 BYTE)	Yes	(null)	12	(null)
13	SITUACION	NUMBER(5,0)	Yes	(null)	13	(null)
14	SITUAC_DES	VARCHAR2(100 BYTE)	Yes	(null)	14	(null)
15	TR_FUENTE	VARCHAR2(100 BYTE)	Yes	(null)	15	(null)
16	TR_FUE_DES	VARCHAR2(100 BYTE)	Yes	(null)	16	(null)
17	LONGITUD	NUMBER	Yes	(null)	17	(null)
18	ESTADOFIS	NUMBER(5,0)	Yes	(null)	18	(null)
19	ESTFIS_DES	VARCHAR2(100 BYTE)	Yes	(null)	19	(null)
20	CIRCULACIO	VARCHAR2(4 BYTE)	Yes	(null)	20	(null)
21	CIRCUL_DES	VARCHAR2(100 BYTE)	Yes	(null)	21	(null)
22	ACCESO	NUMBER(5,0)	Yes	(null)	22	(null)
23	ACCESO_DES	VARCHAR2(100 BYTE)	Yes	(null)	23	(null)
24	COMPETENCI	NUMBER(5,0)	Yes	(null)	24	(null)
25	COMPET_DES	VARCHAR2(100 BYTE)	Yes	(null)	25	(null)
26	ORDEN	NUMBER(5,0)	Yes	(null)	26	(null)
27	ORDEN_DES	VARCHAR2(100 BYTE)	Yes	(null)	27	(null)
28	ALTA_DB	DATE	Yes	(null)	28	(null)
29	GEOM	SDO_GEOMETRY	Yes	(null)	29	(null)

RET. Fuentes. Cartociudad. Portales.

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
ID_POR_PK	NUMBER	Yes	(null)	1	(null)
ID_TRAMO	NUMBER	Yes	(null)	2	(null)
ID_VIAL	NUMBER	Yes	(null)	3	(null)
ID_CP	NUMBER	Yes	(null)	4	(null)
NUM_POR	NUMBER	Yes	(null)	5	(null)
EXTENSION	VARCHAR2(4 BYTE)	Yes	(null)	6	(null)
TIPO_PORTA	NUMBER	Yes	(null)	7	(null)
TIPO_P_DES	VARCHAR2(100 BYTE)	Yes	(null)	8	(null)
FUENTE	VARCHAR2(100 BYTE)	Yes	(null)	9	(null)
FUENTE_DES	VARCHAR2(100 BYTE)	Yes	(null)	10	(null)
ALTA_DB	DATE	Yes	(null)	11	(null)
GEOM	SDO_GEOMETRY	Yes	(null)	12	(null)

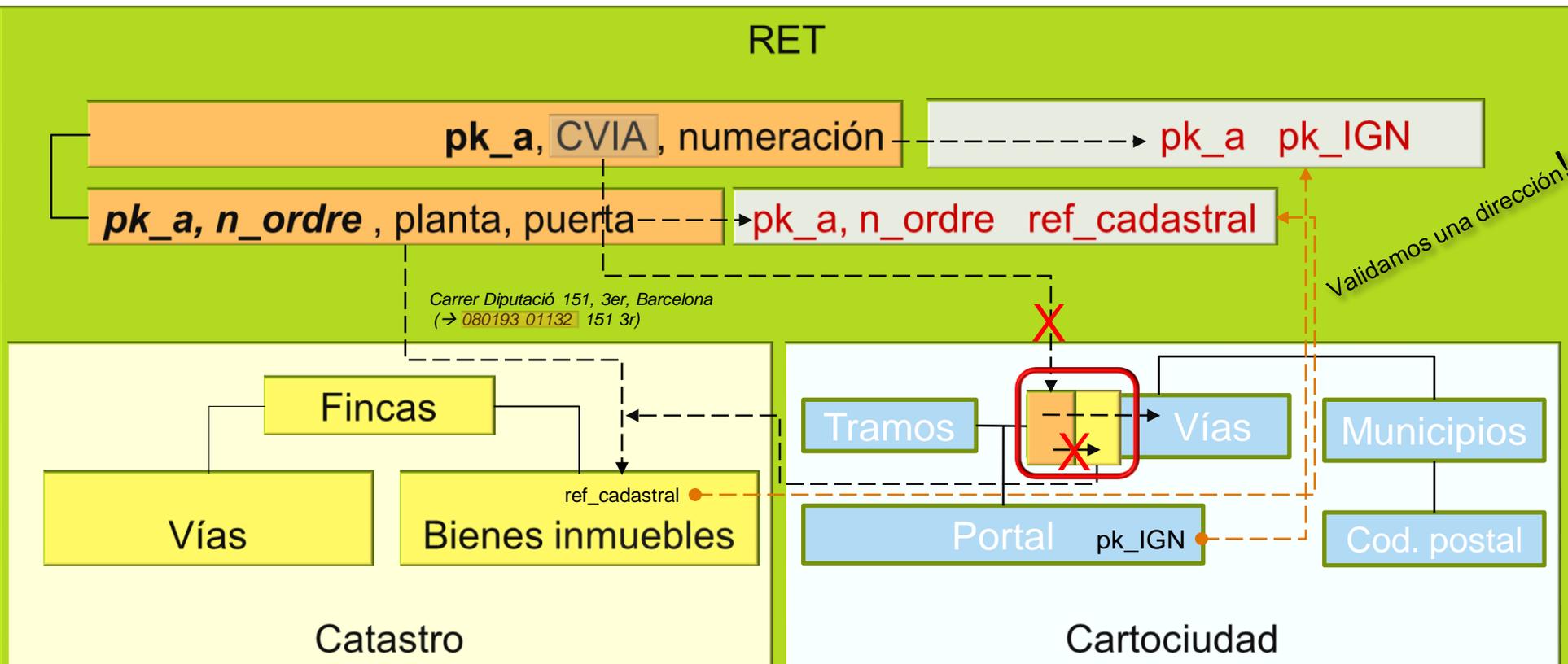
↑
SRID: 4258

-  Portales Cartociudad
-  Portales RP2014



RET. Fuentes. Cartociudad.

- Modeliza direcciones a nivel de portal
- No modeliza direcciones verticales (escalera, planta, puerta)
- Cada portal, tramo i vía tiene una pk. Se establece la relación con vías INE i DGC
- La validación de una dirección vertical se podría realizar vía el archivo CAT del Catastro, en teoría

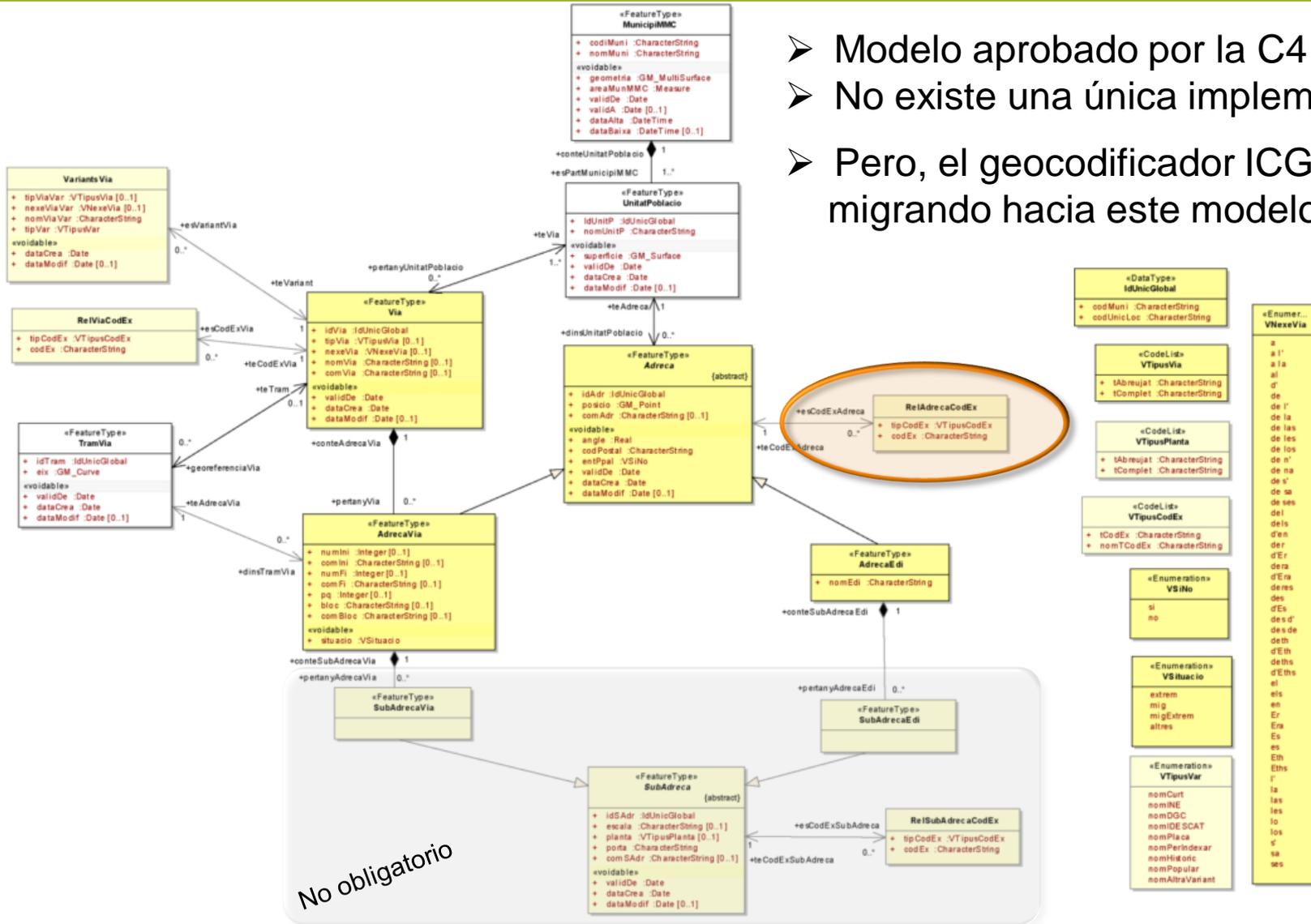


RET. Fuentes.

Fuente	Organismos	Observaciones
Catastro. Archivo CAT	DGC	Modeliza fincas, bienes inmuebles, unidades constructivas etc. <u>Códigos propios, diferentes a INE</u> Fuzzy search.... falsos positivos!!
Cartociudad	IGN (INE;DGC;Correos)	Modeliza vías, tramos de vías, portales, códigos postales
Censo Edificios 2011. Aproximaciones postales.	INE	Contiene información sobre las aproximaciones postales de los edificios incluidos en la <u>operación Censo 2011</u>
Direcciones AGE	INE, IGN, DGC, Correos	Modeliza direcciones. <u>En fase de diseño...</u>
BDMAC	ICGC, municipios, IDESCAT	Modeliza direcciones. Modelo aprobado C4

RET. Fuentes. BDMAC.

- Modelo aprobado por la C4
- No existe una única implementación
- Pero, el geocodificador ICGC está migrando hacia este modelo



RET. Fuentes. Geocodificador ICGC + Catastro.

RET

pk_a, CVIA, TVIA, VIA, numeración

pk_a pk_ICGC, (x,y),

pk_a, n_ordre, planta, puerta

pk_a, n_ordre ref_cadastral_bi,

Calle Diputació 151, 3r, Barcelona
(→ 080193 01132 151 3r)

Validamos una dirección I la relacionamos con otros datos

Vías

ref_cadastral

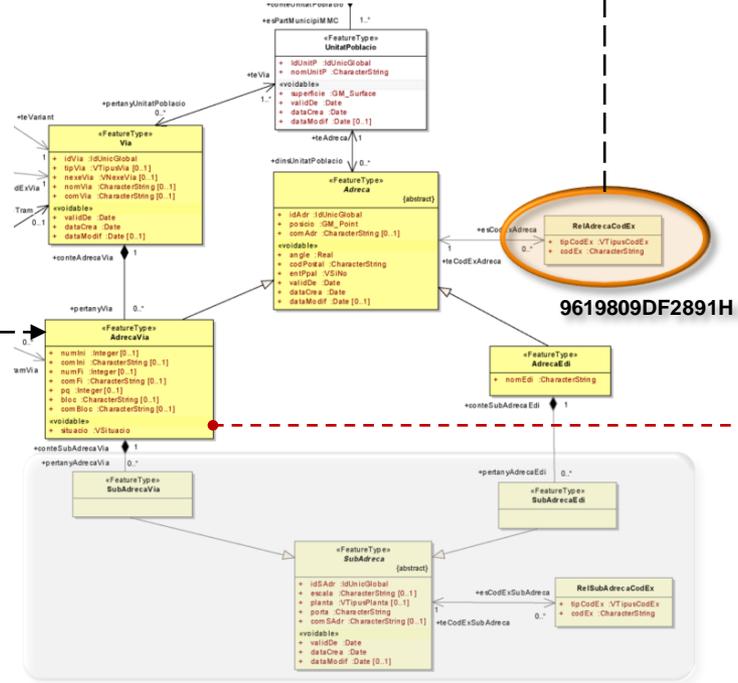
Fincas

Béns immobles

ref_cadastral

parcela_catastral_text	num_ordre_bi_integer	primer_dc_text	segon_dc_text	planta_text	porta_text	any_antiguetat_integer	superficie_bi_integer	clau_grup_text
9619809DF2891H	1	P	P	0		1910	130	C
9619809DF2891H	2	A	A	PR		1910	116	V
9619809DF2891H	3	S	S	01		1910	116	V
9619809DF2891H	4	D	D	02		1910	116	V
9619809DF2891H	5	F	F	03		1910	116	V
9619809DF2891H	6	G	G	04		1910	116	V

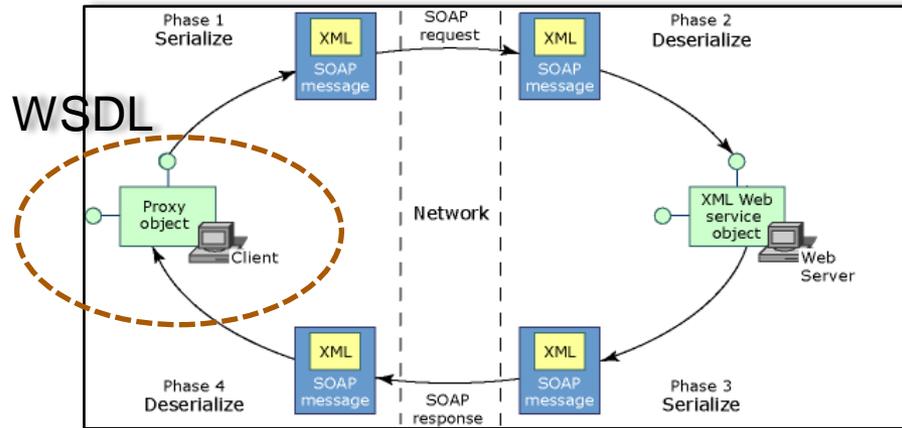
Catastro



Servicio geocodificación ICGC

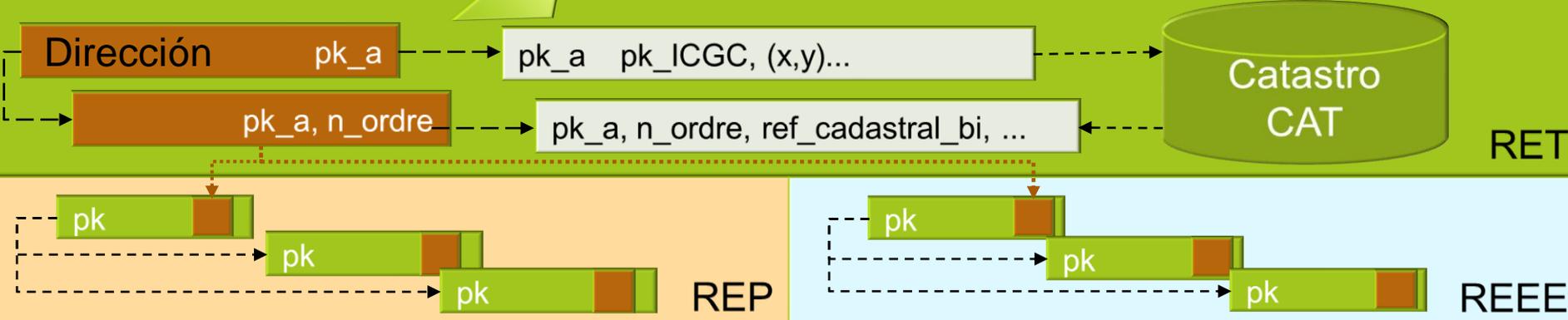
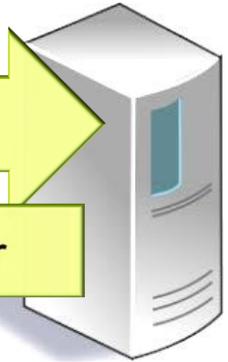
RET. Fuentes. Geocodificador ICGC + Catastro.

Nuevo API SOAP



promun, cvia, tipus_via, nom_via num, fecha

pk_ICGC, ref_cadastral_fi, (x,y),... dsc_error



Gracias



www.idescat.cat



<https://twitter.com/idescat>