

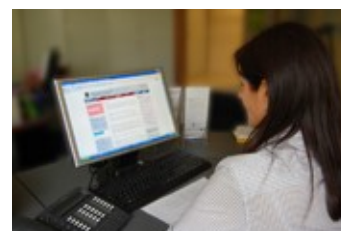
Géolocalisation de fichiers d'adresses

Hackathon champ de sirene

*Jean-François BARON
Division Statistiques et analyses urbaines
Section Partenariats et Géolocalisation*



Mesurer pour comprendre



janvier
2017

Organisation de la géolocalisation

Le but de la géolocalisation :

Attribuer des coordonnées x,y à une adresse et déterminer l'appartenance de cette adresse à des zonages (Iris, QPV, ZFU, *etc.*)

Les outils :

- l'application GEOLOC
- une chaîne de traitements sas hors application

Les équipes :

- A Metz, une équipe géoréférencement qui gère le traitement automatique et des gestionnaires qui font de la reprise manuelle
- A Orléans, un pôle qui s'occupe du contrôle des données et de la diffusion

L'application GEOLOC

- Elle contient un référentiel d'adresses produit par la DMRG, les contours des communes et des zonages.

1) elle tente d'**appairer automatiquement** l'adresse du fichier source (code commune, libellé de voie, numéro de voie) avec une adresse du référentiel pour y récupérer le x,y.

2) en cas d'échec d'appariement, de la **reprise manuelle** peut-être faite par les gestionnaires de Metz (pas le cas pour les fichiers du Hackathon). L'information est conservée dans un fichier de capitalisation qui s'enrichit au fur et à mesure des sources traitées en reprise.

3) le fichier de **capitalisation** est réutilisé pour le traitement des sources suivantes entre l'appariement automatique et la reprise.

Le référentiel d'adresses 1/2

Contient environ 25 millions d'adresses (métropole + 4 dom), uniquement les adresses bâties.

-communes de + 10 000 habitants :

→ les adresses bâties résidentielles (et leur x,y) viennent du RIL

→ les adresses bâties non résidentielles viennent des fichiers fiscaux et leur x,y du plan cadastral. Le lien entre les 2 peut se faire grâce à l'identifiant de parcelle commun aux 2 fichiers. Le x,y de la parcelle est positionné à l'étiquette de parcelle (différent du centroïde)

- communes de – 10 000 habitants :

Toutes les adresses bâties viennent des fichiers fiscaux et leur x,y du plan cadastral

Le référentiel d'adresses 2/2

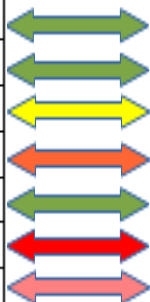
Mise à jour annuelle. Actuellement la géographie est 2015 (idem pour les contours d'Iris et de communes).

Un extrait du référentiel :

Code commune	Type de voie	Libellé de la voie	Numéro dans la voie	x	y	Nombre de logements
<u>73065</u>	BD	BD GAMBETTA	3	<u>927319,60</u>	<u>6501266,10</u>	22
<u>73065</u>	BD	BD GAMBETTA	4	<u>927397,37</u>	<u>6501371,76</u>	0
<u>73065</u>	BD	BD GAMBETTA	5	<u>927284,90</u>	<u>6501271,90</u>	11
<u>73065</u>	BD	BD GAMBETTA	7	<u>927262,19</u>	<u>6501247,25</u>	0
<u>73065</u>	<u>FG</u>	<u>FG MONTMELIAN</u>	15	<u>928198,20</u>	<u>6500522,30</u>	3
<u>73065</u>	<u>FG</u>	<u>FG MONTMELIAN</u>	16	<u>928194,77</u>	<u>6500499,91</u>	0
<u>73065</u>	<u>FG</u>	<u>FG MONTMELIAN</u>	17	<u>928203,30</u>	<u>6500523,30</u>	4
<u>73065</u>	<u>FG</u>	<u>FG MONTMELIAN</u>	31	<u>928216,10</u>	<u>6500525,90</u>	12

Appariement avec le référentiel

Code commune	Type de voie	Libellé de la voie	Numéro dans la voie
<u>73065</u>	BD	BD GAMBETTA	3
<u>73065</u>	BD	BD GAMBETTA	4
<u>73065</u>	BD	BD GAMBETTA	6
<u>73065</u>	R	R GAMBETTA	7
<u>73065</u>	FG	FG MONTMELIAN	15
<u>73065</u>	AV	AV MONT	16
<u>73065</u>	FG	FG MONTMELIAN	39



Code commune	Type de voie	Libellé de la voie	Numéro dans la voie	x	y
<u>73065</u>	BD	BD GAMBETTA	3	<u>927319,60</u>	<u>6501266,10</u>
<u>73065</u>	BD	BD GAMBETTA	4	<u>927397,37</u>	<u>6501371,76</u>
<u>73065</u>	BD	BD GAMBETTA	5	<u>927284,90</u>	<u>6501271,90</u>
<u>73065</u>	BD	BD GAMBETTA	7	<u>927262,19</u>	<u>6501247,25</u>
<u>73065</u>	FG	FG MONTMELIAN	15	<u>928198,20</u>	<u>6500522,30</u>
<u>73065</u>	FG	FG MONTMELIAN	16	<u>928194,77</u>	<u>6500499,91</u>
<u>73065</u>	FG	FG MONTMELIAN	17	<u>928203,30</u>	<u>6500523,30</u>
<u>73065</u>	FG	FG MONTMELIAN	31	<u>928216,10</u>	<u>6500525,90</u>

Appariement de l'adresse source avec une adresse du référentiel

=> affectation du x,y du référentiel à l'adresse source

La voie source est trouvée dans le référentiel mais pas l'adresse

=> interpolation du x,y car le 5 et le 7 sont dans le référentiel

La voie source est trouvée dans le référentiel mais il existe un doute

=> passage en capitalisation. Si adresse trouvée Geoloc renvoie généralement un x,y sinon Geoloc renvoie tous les x,y de la voie

La voie source est trouvée dans le référentiel mais pas l'adresse. Interpolation impossible.

=> passage en capitalisation. Si adresse trouvée Geoloc renvoie généralement un x,y sinon Geoloc renvoie tous les x,y de la voie

La voie source n'est pas trouvée dans le référentiel.

=> passage en capitalisation. Si adresse trouvée Geoloc renvoie généralement un x,y sinon Geoloc ne renvoie rien.

Chaîne de traitement sas hors application

En sortie de GEOLOC une adresse peut se retrouver dans une des 3 situations suivantes :

- 1) elle a **un seul x,y** et ses zonages lui ont été affectés
- 2) elle a **plusieurs x,y** et certains zonages ont pu lui être attribués (par exemple si tous les x,y se trouvent dans le même Iris), d'autres non.
- 3) elle n'a **aucun x,y** et donc peu de chance d'avoir un zonage

Les adresses dans les situations 2 et 3 vont être traitées hors application dans les « post traitements ». S'il y avait eu reprise manuelle, elle aurait eu lieu avant les post traitements.

Remarque : Certaines adresses du fichier source ne passent même pas dans GEOLOC quand le code commune n'est pas considéré comme valide (étranger, mauvaise géographie, etc.). C'est le cas de 9 % des adresses employeurs.

Imputation de x,y et injection de zonage

But de la chaîne de traitement hors application : imputer un x,y aux adresses dans les situations 2 et 3 puis injecter les zonages à partir du x,y.

Situation 2 : Regrouper les x,y proposés par GEOLOC par carreau de 100 mètres de côté, leur affecter le x,y du centre du carreau et la somme du nombre de logements des adresses correspondant à ces x,y. Le x,y affecté à l'adresse est un centre de carreau imputé aléatoirement en pondérant par le nombre de logements.

Situation 3 : Affecter à l'adresse un x,y centre d'un carreau de la commune par imputation aléatoire pondérée par le nombre de logements du carreau

Injection du zonage à partir du x,y

Qualité de l'appariement

Qual est un indicateur de la qualité de l'appariement de l'adresse du fichier source avec une adresse du référentiel :

1 : adresse source trouvée dans le référentiel ou par capitalisation
=> x,y du référentiel attribué

2.1 : adresse source non trouvée dans le référentiel mais 2 adresses encadrantes de même parité trouvées dans la même voie (par exemple : 3 encadré par 1 et 5) => x,y interpolé linéairement

2.2 : adresse source non trouvée dans le référentiel mais 2 adresses encadrantes trouvées de parité inverse dans la même voie (par exemple : 3 encadré par 2 et 4) => x,y interpolé linéairement

8 : adresse non trouvée mais voie sûre => x,y imputé à la voie

8.5 : adresse non trouvée et voie douteuse => x,y imputé à la voie

9 : adresse et voie non trouvées => x,y imputé à la commune

Qualité des adresses employeurs du RP

qual	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1	501316	32.82	501316	32.82
2.1	49607	3.25	550923	36.07
2.2	19263	1.26	570186	37.33
8	216427	14.17	786613	51.49
8.5	137039	8.97	923652	60.47
9	603923	39.53	1527575	100.00
Fréquence manquante = 146389				

Erreur potentielle

Sourcexyw : distance maximale entre le x,y imputé et le x,y réel.

- quand il y imputation à la voie
- sous réserve que le référentiel soit juste

Merci de votre attention !