

# La géolocalisation du Recensement de la population dans les communes métropolitaines de moins de 10 000 habitants

JMS 2022



*Travaux réalisés dans le cadre d'un Grant Eurostat*

**1 GÉNÉRALITÉS ET CONTEXTE**

**2 DEUX MÉTHODES DE GÉOLOCALISATION**

**3 MESURE DE LA QUALITÉ**

**4 CHOIX DES COORDONNÉES**

# 01 GÉNÉRALITÉS ET CONTEXTE

En métropole, la méthode de recensement dépend de la taille de la commune et de la catégorie de population

Elle s'appuie sur un cycle de 5 Enquêtes annuelles de recensement (EAR)

	Logements ordinaires			
	Communes de 10 000 hab. et plus (grandes communes)	Communes de moins de 10 000 hab. (petites communes)	Communautés	Habitations mobiles et sans-abri
Recensement de la population	Sondage	Exhaustif au cours d'un cycle de 5 ans		

Les petites communes sont recensées – par une EAR – exhaustivement tous les 5 ans, de façon tournante.

**Le Recensement européen 2021** : fourniture de données sur des carreaux de 1km<sup>2</sup>

**Enjeu** : disposer d'une géolocalisation pour toutes les adresses

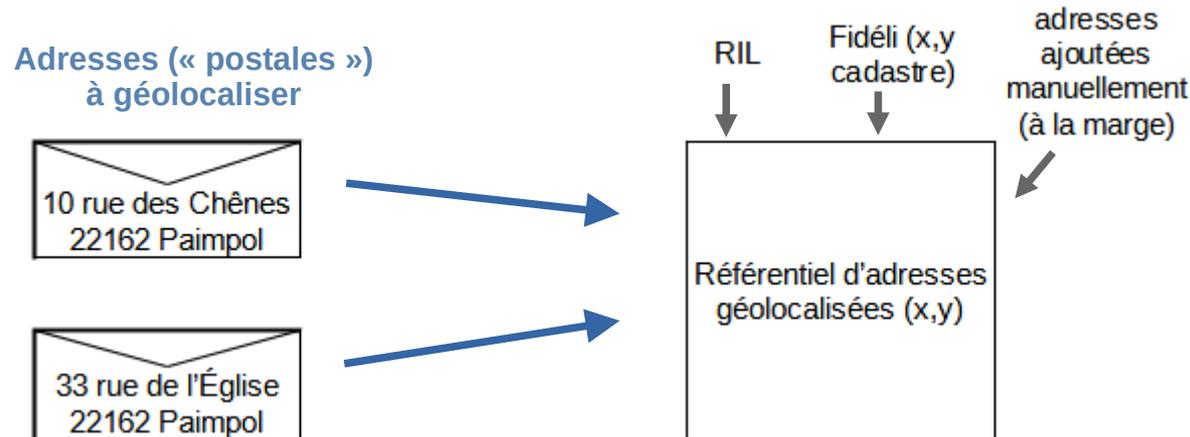
Logements ordinaires

	Communes de 10 000 hab. et plus (grandes communes)	Communes de moins de 10 000 hab. (petites communes)	Communautés	Habitations mobiles et sans-abri
Recensement de la population	Par sondage	Exhaustif au cours d'un cycle de 5 ans		
Adresses géolocalisées	Oui, <b>via le RIL</b>	<b>Pas de géolocalisation de référence</b>	Oui, <b>via le répertoire des communautés</b>	Non (non demandé par Eurostat)

# 02 DEUX MÉTHODES DE GÉOLOCALISATION

- Géolocalisation à partir des **éléments d'adressage** - méthode GA (numéro de voie, nom de voie, code commune, lieu-dit), mise en œuvre par l'application **Géoloc**

## Principe :



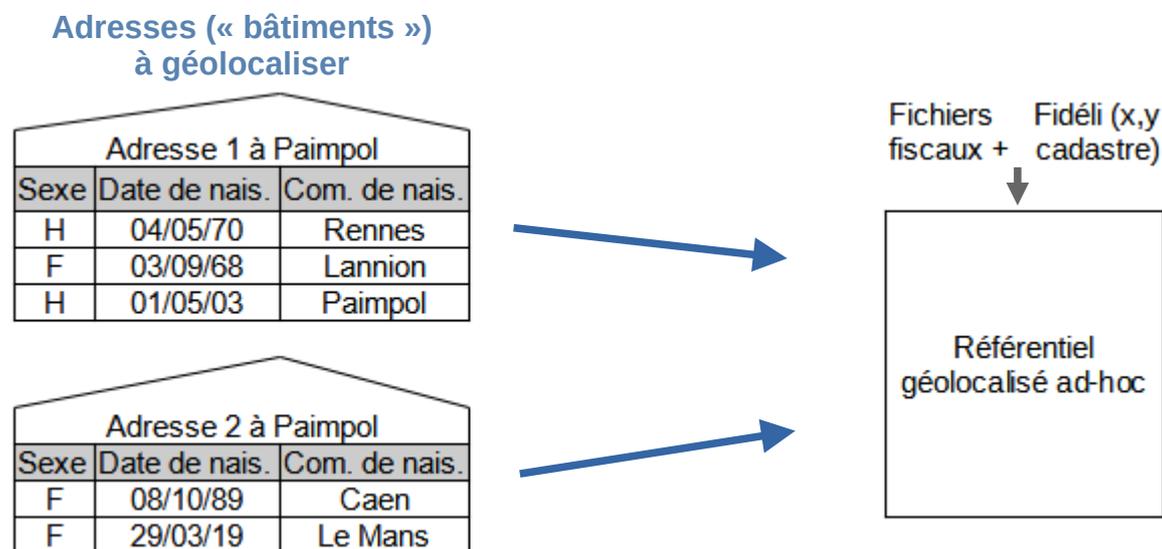
- Processus standard pour la géolocalisation des fichiers d'adresses à l'Insee. Service proposé aux SSM et péri SSP (allocataires de l'assurance maladie, etc.)

La méthode GA produit une **note qualité** : exemple sur les petites communes de l'EAR 2017

<b>Note qualité initiale méthode GA</b>	<b>Nombre d'individus</b>	<b>Part d'individus</b>
11. Voie Sûre, Numéro trouvé	5 019 575	76,9 %
12. Voie Sûre, Position aléatoire dans la voie	492 929	7,6 %
21. Voie probable, Numéro trouvé	358 762	5,5 %
22. Voie probable, Position aléatoire dans la voie	265 438	4,1 %
33. Voie inconnue, Position aléatoire dans la commune	389 278	6,0 %
<b>TOTAL</b>	<b>6 525 982</b>	<b>100,0 %</b>

- Géolocalisation à partir des **caractéristiques des individus** - méthode GI (commune de résidence, sexe, date de naissance, lieu de naissance)

## Principe :



- **Méthode développée** pour calculer des indicateurs au niveau des quartiers prioritaires de la politique de la ville (QPV)

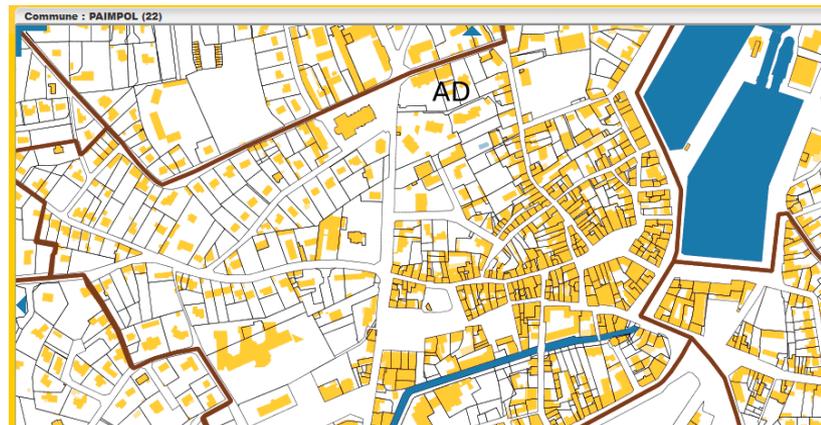
Un système de notation initial qui se base uniquement sur la qualité de la coordonnée dans Fidéli (ne rend pas compte de la qualité de l'appariement avec le référentiel)

Répartition des notes initiales sur les petites communes de l'EAR 2017 :

Note qualité initiale méthode GI	Nombre d'individus	Part d'individus
1. Apparié avec le référentiel, les coordonnées sont de bonne qualité	6 234 465	95,5 %
2. Apparié avec le référentiel, les coordonnées sont de qualité moyenne	7 875	0,1 %
3. Apparié avec le référentiel, les coordonnées sont de mauvaise qualité	3 708	0,1 %
4. Non apparié (imputé)	231 208	3,5 %
5. Non apparié (pas de coordonnée)	48 726	0,7 %
<b>TOTAL</b>	<b>6 525 982</b>	<b>100,0 %</b>

## Les méthodes GI et GA utilisent :

- des coordonnées provenant le plus souvent de la même source (cadastre via Fideli)
- une méthode d'appariement qui se base sur des critères complètement différents



## Comparaison des deux méthodes sur l'EAR 2017 :

- Dans 57 % des cas les coordonnées fournies par les deux méthodes sont identiques, dans 82 % des cas elles sont à moins de 100 mètres
- En moyenne, elles sont éloignées de 157 mètres
- Dans 90 % des cas elles sont dans le même carreau Eurostat, dans 8 % des cas dans un carreau limitrophe

# 03

## MESURE DE LA QUALITÉ

---

On dispose de deux méthodes distinctes de géolocalisation pour les petites communes. Pour chaque adresse, on veut en choisir une :

- Soit ses coordonnées sont identiques : peu importe laquelle on choisit
- Soit ses coordonnées sont différentes : comment savoir laquelle est la bonne ? En particulier, quand les deux méthodes renvoient une note de bonne qualité (cas majoritaire)

**Objectif : mieux cibler les très bonnes notes** i.e. identifier des groupes parmi les notes “bonnes” de chaque méthode qui seraient de qualité moindre

→ Pour ce travail, on se base sur la distance entre les coordonnées fournies par les deux méthodes pour une adresse

		Note qualité initiale méthode GI					
<i>Nombre d'individus</i>		1. Apparié avec le référentiel, les coordonnées sont de bonne qualité	2. Apparié avec le référentiel, les coordonnées sont de qualité moyenne	3. Apparié avec le référentiel, les coordonnées sont de mauvaise qualité	4. Non apparié (imputé)	5. Non apparié (pas de coordonnée)	TOTAL
Distance moyenne entre coordonnées GI et GA (en m)							
<b>Note qualité initiale Méthode GA</b>	<b>11. Voie Sûre, Numéro trouvé</b>	4 821 704	5 499	660	162 136	29 576	5 019 575
		20 m	68 m	986 m	155 m		24 m
	<b>12. Voie Sûre, Position aléatoire dans la voie</b>	462 563	1 018	987	22 753	5 608	429 929
		242 m	232 m	663 m	584 m		256 m
	<b>21. Voie probable, Numéro trouvé</b>	338 754	520	213	13 231	6 044	358 762
		128 m	224 m	1 140 m	549 m		142 m
	<b>22. Voie probable, Position aléatoire dans la voie</b>	248 916	119	423	12 040	3 940	265 438
		302 m	372 m	739 m	820 m		322 m
	<b>33. Voie inconnue, Position aléatoire dans la commune</b>	362 528	719	1 425	21 048	3 558	389 278
		1 674 m	1 509 m	1 441 m	1 527 m		1 649 m
	<b>TOTAL</b>	6 234 465	7 875	3 708	231 208	48 726	6 525 982
		149 m	236 m	1 055 m	380 m		157 m

- Avec GA, lorsqu'on cherche à géolocaliser une adresse **sans numéro de voie**, elle peut obtenir la note maximale.
  - En pratique la coordonnée est plus imprécise dans ces cas
- On scinde la note « 11. Voie sure, numéro trouvé » en deux groupes :
  - «11a. Voie sure, numéro trouvé, avec numéro » (ex : 12, rue de l'Église)  
**=> Qualité : Bonne**
  - «11a. Voie sure, numéro trouvé, sans numéro » (ex : rue de l'Église)  
**=> Qualité : Moyenne**
- On regroupe les autres notes : Moyen (11b +21), Mauvais (12+22)

Ce faisant, on cible un peu mieux les “bonnes” coordonnées GA

	Nombre d'individus Distance moyenne entre coordonnées GI et GA (en m)	Note qualité initiale méthode GI					TOTAL
		1. Apparié avec le référentiel, les coordonnées sont de bonne qualité	2. Apparié avec le référentiel, les coordonnées sont de qualité moyenne	3. Apparié avec le référentiel, les coordonnées sont de mauvaise qualité	4. Non apparié (imputé)	5. Non apparié (pas de coordonnée)	
<b>Nouvelle note synthétique Méthode GA</b>	<b>1. Bon</b>	4 761 026 18 m	5 482 68 m	557 994 m	159 478 144 m	28 160	4 954 703 22 m
	<b>2. Moyen</b>	399 432 129 m	537 224 m	316 1 074 m	15 889 601 m	7 460	423 634 146 m
	<b>3. Mauvais</b>	711 479 263 m	1 137 247 m	1 410 685 m	34 783 666 m	9 548	758 367 279 m
	<b>4. Non apparié</b>	362 538 1 674 m	719 1 509 m	1 425 1 441 m	21 048 1 527 m	3 558	389 278 1 649 m
	<b>TOTAL</b>	6 234 465 149 m	7 875 236 m	3 708 1 055 m	231 208 380 m	48 726	6 525 982 157 m

**Des adresses ont vu leur note GA rétrogradée (violet -> rose) :**

- Avant on ne savait pas quelle coordonnée choisir pour ces adresses
- Maintenant, on sait qu'il faut choisir la coordonnée issue de la méthode GI



- Ce faisant, on cible un peu mieux les “bonnes” coordonnées issues de GI

		Nouvelle note synthétique Méthode GI					
Nombre d'individus		1. Bon	2. Moyen	3. Mauvais	4. Non apparié (imputé)	5. Non apparié (pas de coordonnée)	TOTAL
Distance moyenne entre coordonnées GI et GA (en m)							
Nouvelle note synthétique Méthode GA	1. Bon	3 597 700 12 m	1 044 570 24 m	124 795 149 m	159 478 144 m	28 160	4 954 703 22 m
	2. Moyen	299 710 116 m	89 729 141 m	10 846 432 m	15 889 601 m	7 460	423 634 146 m
	3. Mauvais	532 207 248 m	159 960 274 m	21 859 558 m	34 783 666 m	9 548	758 367 279 m
	4. Non apparié	267 299 1 676 m	85 693 1 658 m	11 680 1 693 m	21 048 1 527 m	3 558	389 278 1 649 m
	TOTAL	4 696 919 140 m	1 379 952 162 m	169 180 326 m	231 208 380 m	48 726	6 525 982 157 m

Des adresses ont vu leur note GI rétrogradée :

- Avant on ne savait pas quelle coordonnées choisir pour ces adresses
- Maintenant on sait qu'il faut choisir la coordonnée de la méthode GA

# 04 CHOIX DES COORDONNÉES

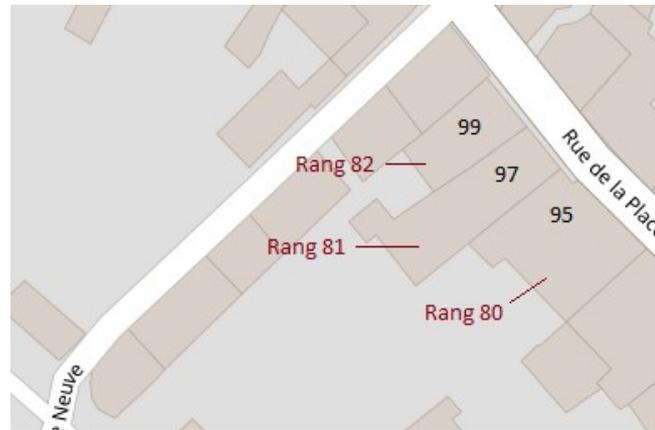
- **Règle de décision :**
  - Lorsque les coordonnées GA sont bonnes, on les choisit : 76 % des individus
  - Sinon, lorsque les coordonnées GI sont bonnes, on les choisit : 16,8 % des individus
  - Sinon, lorsque les coordonnées GI sont moyennes, on les choisit : 5,1 % des individus
  - Sinon, lorsque les coordonnées GA sont moyennes, on les choisit : 0,5 % des individus
  - Sinon, on interpole : 1,6 % des individus
- **Le choix de GA quand les deux notes sont bonnes n'est déterminant que pour une petite partie des adresses**

Quand les deux méthodes fournissent une coordonnée de bonne qualité (nouvelle note) :

  - dans 85,2 % des cas où les deux méthodes fournissent une coordonnée de bonne qualité, les coordonnées proposées par les deux méthodes sont distantes de moins de 10 mètres
  - dans 98,2 % des cas les coordonnées sont distantes de moins de 100 mètres.

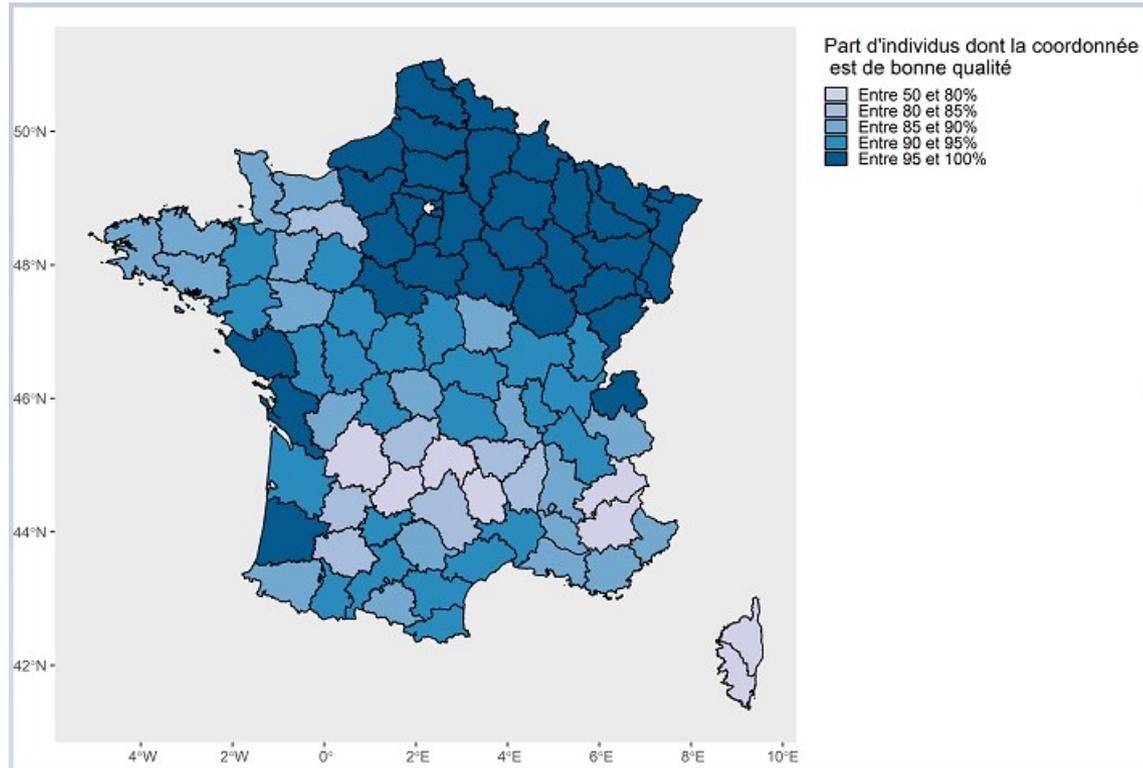
## Quand aucune méthode ne fournit de coordonnée satisfaisante, on interpole :

- Rangs d'adresse : avant la collecte, les agents recenseurs numérotent les adresses qu'ils vont enquêter, dans l'ordre de leur tournée
- Le plus souvent, les rangs d'adresse consécutifs correspondent à des adresses adjacentes



- Le rang d'adresse 81 est calculé comme le barycentre des rangs d'adresses 80 et 82
- Même si les coordonnées des adresses adjacentes ne sont pas connues, on interpole avec les plus proches adresses

## Part des coordonnées de bonne qualité (GA ou GI), par département



Dans certains territoires, coordonnées de moindre qualité :

- C'est surtout la part de coordonnées GA de bonne qualité qui est variable selon les territoires, en raison des différences dans la normalisation des adresses
- Habitat peu dense
- Pas d'obligation d'avoir des noms de voirie dans les communes de moins de 2000 habitants

## Améliorations envisagées :

- Possibilité d'**assouplir les conditions d'appariement** de la méthode GI pour les rendre moins binaires ?
- Pour les communes qui ont plusieurs districts de collecte, contrôler les coordonnées aberrantes en **comparant les distances au sein d'un district**

EAR

Adresse 3 à Rocamadour		
Sexe	Date de naissance	Lieu de naissance
F	01/08/1990	Cahors

0 individu  
retrouvé

Référentiel ad-hoc

Adresse D à Rocamadour		
Sexe	Date de naissance	Lieu de naissance
F	07/08/1990	Cahors



Commune de Paimpol, district 0056,  
exemple de voie probablement mal localisée



× Adresse du district 0056 de la commune

## Retrouvez-nous sur

[insee.fr](https://www.insee.fr)



---

Jeanne PAGÈS (\*) et Gabrielle GALLIC (\*\*)

(\*) *Insee, Département de la Démographie*

(\*\*) *Insee, Département de l'Action Régionale*