RETOUR D'EXPÉRIENCE MOBILISER DES DONNÉES DE GESTION ADMINISTRATIVE LOCALISÉES (CAF) POUR LA RECHERCHE EN GÉOGRAPHIE

Luc Guibard
Université Paris Diderot- UMR Géographie-cités

Journée d'étude

Données individuelles localisées: enjeux juridiques, éthiques et technique à l'heure du RGPD CIST – 15 novembre 2019



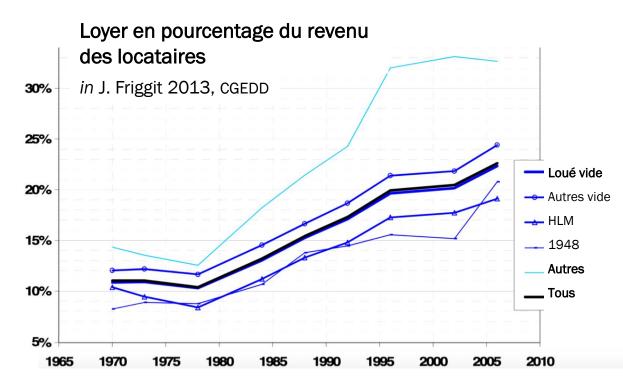




« Les dynamiques spatiales de la gentrification à Paris » Fronts de Avant-postes de gentrification gentrification Noyau historique des beaux-quartiers 1970 Forte concentration actuelle de population étrangère Couronne de logements sociaux

in A. Clerval, 2010

2000





L. Guibard, Le Plessis-Robinson, 2019

Quantifier l'éviction - Le problème du suivi des mobilités résidentielles

- Peu de jeu de données permettent un suivi précis des mobilités résidentielles (fichiers Migcom de l'Insee).
- « Mesurer l'invisible » (Atkinson 2000)
- Les travaux sur la gentrification ont été principalement basés sur la comparaison d'effectifs de population à différentes dates.
- Des exceptions: Wyly 2010, Petsimeris 2015, Giroud et Ter Minassian 2016.

La numérisation des données administratives offre de nouvelles possibilités

Flux vers Paris

Limites territoriales

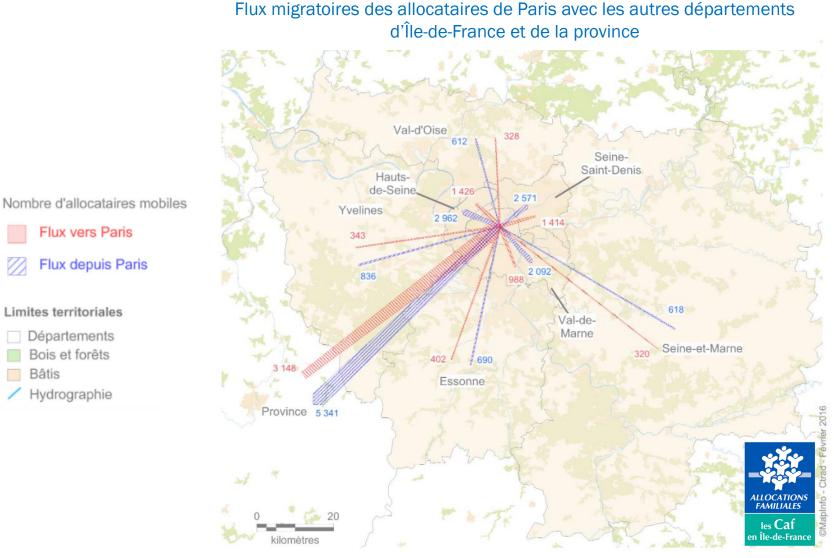
Départements

/ Hydrographie

Bâtis

Bois et forêts

Flux depuis Paris



in Chemineau D., Bergeron T., 2016, « La mobilité résidentielle des allocataires parisiens », Dossier d'études des CAF en Ile-de-France n°10.

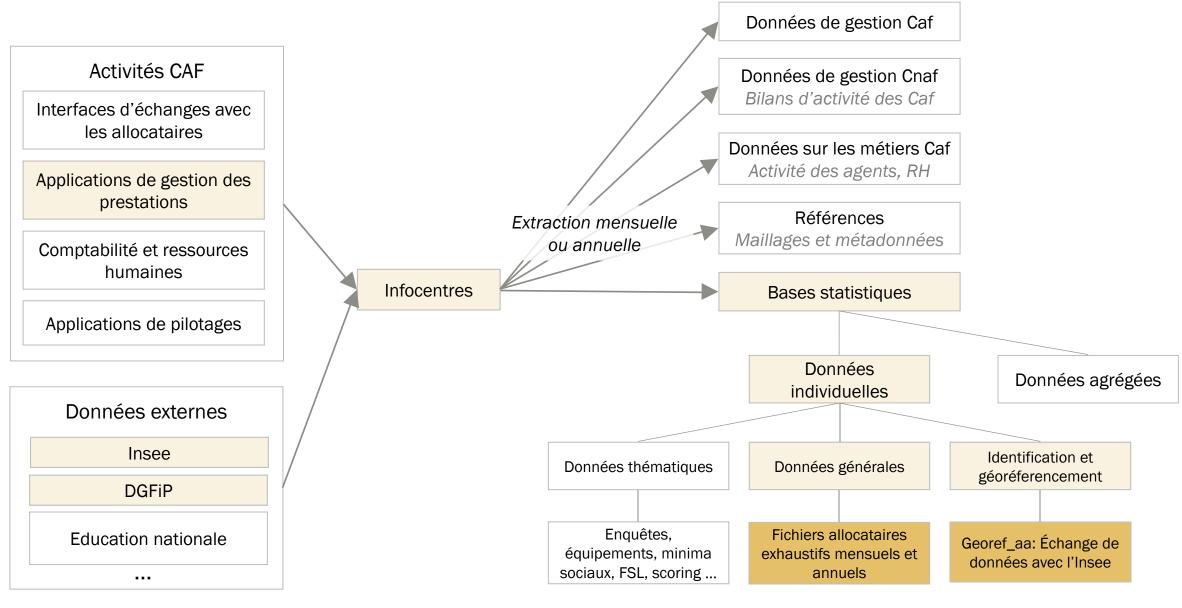
Source statistique : Caisses d'allocations familiales d'Île-de-France, fichiers réduits exhaustifs mensuels au 30 novembre 2012 et au 30 novembre 2013.

LES DONNÉES CAF

Obtenir l'accès aux données

- Cadre juridique: convention de mise à disposition des données (Caf du Val de Marne, Université Paris-Diderot).
- Accueil au sein de la Cellule Technique de Réflexion et d'Aide à la décision des Cafs d'Ile-de-France (CTRAD)
- Accès sur place aux serveurs de la Caf du Val-de-Marne, via SAS et un portail sécurisé.
- Un processus compliqué.

LES DONNÉES CAF



LES DONNÉES CAF

Données de gestion Caf

Données de gestion Cnaf Bilans d'activité des Caf

Donnees externes

Une source d'information très riche: fréquence d'actualisation, changements d'adresses et géolocalisation, données DGFiP.

Extraction mensuelle

- À caractère personnel au sens du RGPD: individus identifiables directement.

Contenant des informations particulièrement sensibles: revenus, adresses, composition des foyers.

- Produites par une administration publique, elles sont soumises au secret statistique.

Applications de pilotages

etes, Fichiers allocataires
ts, minima exhaustifs mensuels

Georef_aa: Echange de données avec l'Insee

LES DONNÉES CAF - Assurer l'anonymat des données

D'après la convention:

- Aucun dénombrement inférieur à cinq.
- Lois de 1951 et 1978 : aucun renseignement individuel, ni aucun code, numéro de matricule ou moyens de décodage permettant d'identifier directement ou indirectement une personne.
- Ne pas utiliser les données à des fins autres que celles spécifiées dans la convention.
- Des critères qui restent imprécis pour la protection des données contre l'identification indirecte.
- Pas de prise en compte des méthodes statistiques plus poussées de gestion de la confidentialité (lissage, imputation).
- Des correspondants Informatique et Libertés au sein de la Caf du Val de Marne et de la Cnaf.

Population allocataire et population générale en lle-de-France en 2015

		Données allocataire	Insee et DGFiP
Population (millions)		5,83 (48,3 %)	12,08
Taille moyenne des ménages		2,8 (foyer allocataire)	1,78 (foyer fiscal)
			2,3 (résidence principale)
Population de moins de 20 ans		2,73 millions	3,11 millions
		47 % de la population allocataire	26% de la population régionale
		88 % des franciliens de moins de 20 ans	
Composition des ménages	Personne isolée	34,0 %	36,2 %
	Famille monoparentale	16,5 %	10,7 %
	Couple avec enfants	45,2 %	28,8 %
	Couple sans enfants	4,3 %	21,0 %
Part de la population sous le seuil de pauvreté		28,6% (seuil de bas revenus)	15,9 % (seuil de pauvreté)

Source : Cnaf-Dser Allstat FR2_1214, Insee RP2015 et DGFiP 2015

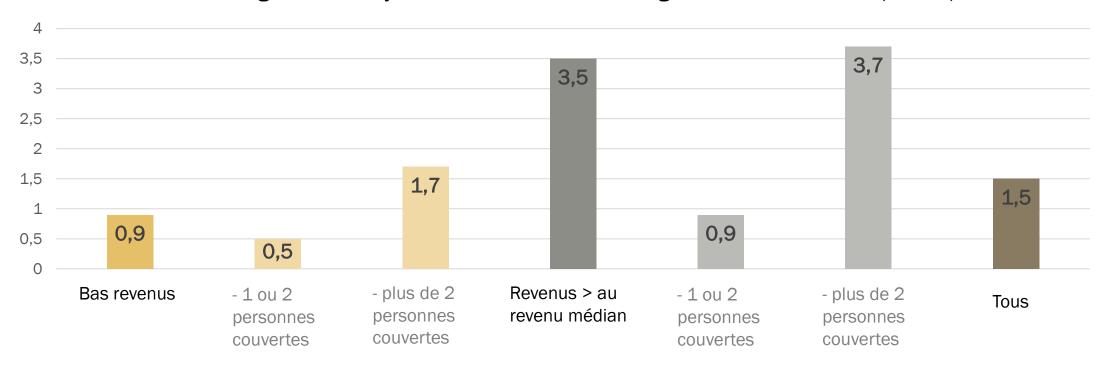
LES DONNÉES CAF - Difficultés rencontrées dans l'utilisation des données

- Comparabilité avec les données du recensement.
- Comprendre le fonctionnement de la base de données.
- A chaque prestation sa population et ses variables. Les variables transversales ne sont pas toujours renseignées de manière homogène (revenu) ou fiable (activité).
- Des données de gestion : niveaux de consolidation et mutations des dossiers.
- Tous les individus n'ont pas le même statut.

Apprentissage au contact des chargés d'études de la CTRAD

A. Indicateurs non localisés

DÉMENAGEMENTS EN ILE-DE-FRANCE EN 2017 Éloignement moyen de Paris lors d'un changement de commune (en km)



N = 181 308 Cnaf-Dser Allstat FR2_1216 et FR2_1217

B. Agrégations sur maillage ad hoc

Exemple 1 : Les ménages déplacés par la rénovation urbaine du Plessis-Robinson

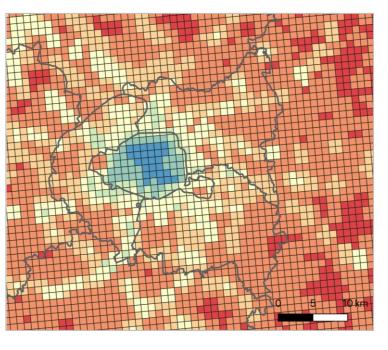






L. Guibard, Le Plessis-Robinson 2015 et 2019

Exemple 2: Niveau de desserte en transports en commun (rayon de 1250 m depuis le centre du carreau)





Aucune

1-3

4 -

8 - 12

13 - 21

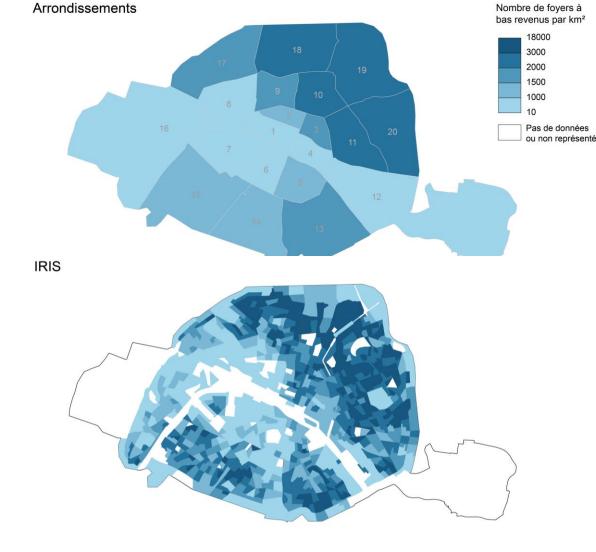
22 - 36

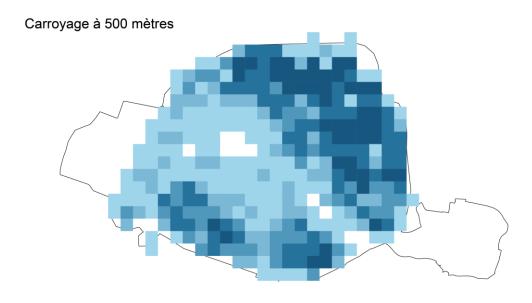
Sources:

OSM (bus et transiliens), sept. 2019 STIF (métro, RER, tram), 2019. INSEE (carroyage), 2019.

L. Guibard, 2019

C. Agrégation à la commune, l'IRIS ou au carreau (avec seuil de confidentialité)

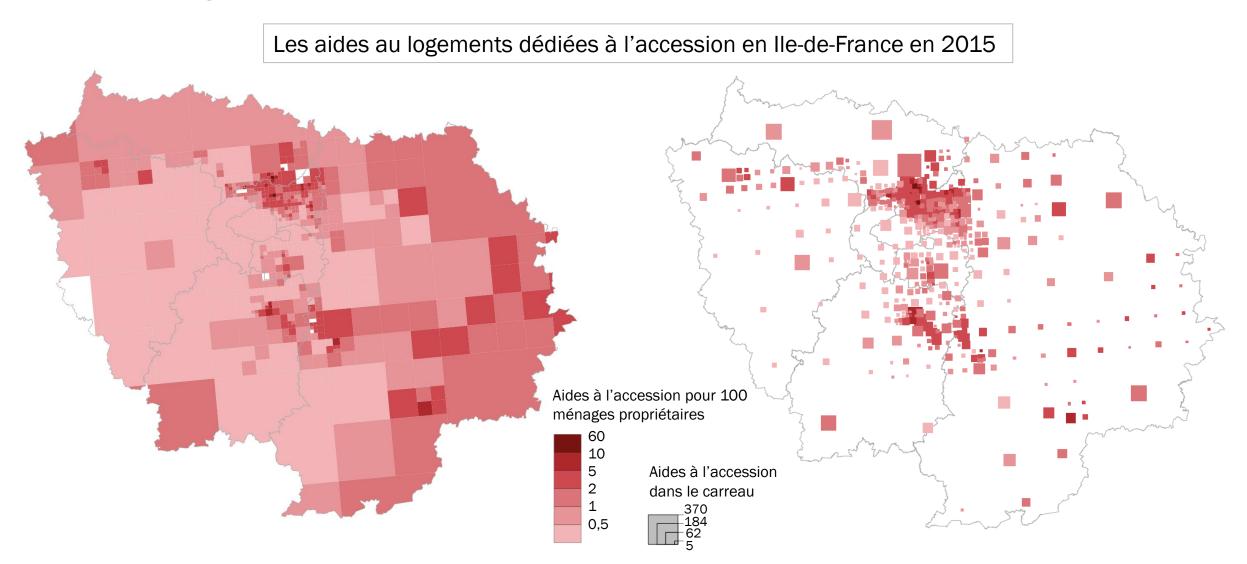




Données: Caisses d'allocations familiales d'Ile-de-France (2016) Auteurs: Luc Guibard (Univ. Paris Diderot), Romain Leconte (Univ. Paris Diderot) et Bruno Fayard (CTRAD), 2018

- Communes: niveau peu précis mais faible perte de données. Échelle privilégiée des politiques urbaines.
- Les IRIS prennent en compte les ruptures morphologiques.
- Problème du MAUP (S. Openshaw, 1984)

D. Carroyage au niveau naturel



Source: Cnaf-Dser, allstat FR2_1215, Insee Filosofi 2015

Auteur: Luc Guibard 2019

Conclusion

- La géolocalisation multiplie les possibilités d'analyse mais elle complexifie les enjeux liés la confidentialité des données, en particulier lorsqu'elle permet les croisements avec d'autres bases.
- Une démarche par tâtonnements : la convention fournit des critères censés garantir
 l'anonymat des données, mais ces derniers doivent être précisés, en particulier en ce qui concerne les risques d'identification indirecte.
- Des méthodes de gestion de la confidentialité à expérimenter: imputation, lissage (potentiel Stewart).
- L'accès aux données a été compliqué, mais l'écriture de la convention devrait faciliter de nouvelles demandes.

"Les Caf utilisent le « datamining » en matière de maîtrise des risques financiers (lutte contre la fraude, vérification des dossiers, contrôle interne...). Le datamining est un processus d'extractions automatiques d'informations qui est semblable au profilage. Cette méthode basée sur des modèles statistiques, a permis de déterminer des «modèles à risque». Quand des dossiers présentent des caractéristiques plus ou moins proches de ces modèles, la Caf leur attribue un «score de risque». Les dossiers ciblés font ainsi l'objet d'un classement en fonction du niveau de risque. Grâce au datamining, les Caf ont gagné en efficience".

Caf du Loiret, lettre au partenaire, Août 2018.