

Note d'analyse n° 133

« [Quinze ans de PNRU : quels effets sur l'habitat et le peuplement ?](#) »¹

ANNEXE MÉTHODOLOGIQUE

Constitution de la base harmonisée des revenus localisés 1999-2019

par Clément Dherbécourt et Gabrielle Jacob

Ce document présente les différentes étapes de la constitution de la *Base harmonisée des revenus localisés 1999-2019*. Cette base est utilisée par France Stratégie dans le cadre de l'évaluation des effets du Programme national de rénovation urbaine (PNRU).

1. De l'intérêt de disposer d'une base harmonisée de revenus localisés sur vingt ans

Afin d'analyser les changements de composition sociale des quartiers, processus progressif et lent par nature, il est nécessaire de disposer de données de long terme par quartier. France Stratégie a publié en 2020 des analyses sur la ségrégation résidentielle à partir des données de recensement harmonisé SAPHIR 1990-2015². L'étude de 2020 se focalise sur les agglomérations de plus de 100 000 habitants, qui ont été découpées en quartiers aux contours stables dans le temps³. Les individus ont été distingués selon la profession des individus, leur ascendance migratoire, leur âge et le fait d'occuper ou non un logement social.

¹ Cusset P.-Y., Dherbécourt C., Guyon N. et Jacob G. (2024), « [Quinze ans de PNRU : quels effets sur l'habitat et le peuplement ?](#) », *La Note d'analyse*, n° 133, France Stratégie, décembre.

² Botton H., Cusset P.-Y., Dherbécourt C. et Alban G. (2020), « Quelle évolution de la ségrégation résidentielle en France ? », France Stratégie, *La Note d'analyse*, n° 92, juin 2020.

³ Les quartiers correspondent aux IRIS, soit des ensembles cohérents du point de vue du bâti et qui comptent en moyenne 2000 habitants. Les quartiers sont harmonisés pour tenir compte des scissions et absorptions entre IRIS ayant eu lieu entre 1999 et 2015. Pour l'année 1990, les quartiers ont été reconstitués à partir des îlots, une unité géographique plus fine que l'IRIS. Voir Botton *et al.* (2020), « L'évolution de la ségrégation résidentielle en France : 1990-2015 », document de travail, France Stratégie, juin, pour plus de détails.

L'analyse met en évidence la stabilité des indices de ségrégation des différentes catégories de population entre 1990 et 2015, à l'exception des locataires de logements sociaux, dont la répartition entre quartiers s'homogénéise au cours du temps.

L'étude de 2020 admet plusieurs limites méthodologiques. D'une part, elle repose sur des catégories sociales dont le poids et les caractéristiques sont variables dans le temps et dans l'espace. D'autre part, les indices de ségrégation sont mesurés pour chaque catégorie (ségrégation des cadres par rapport au reste de la population en âge de travailler, ségrégation des immigrés non-européens par rapport à l'ensemble de la population), sans qu'on puisse en tirer un indice synthétique de ségrégation globale à l'échelle d'une agglomération.

L'approche par groupes de revenus permet de résoudre ces deux difficultés. Des analyses sur la répartition des groupes de revenus entre quartiers ont été utilisées à France Stratégie dans une étude publiée en 2021 sur l'effet de la loi SRU, portant sur les années 2012-2018⁴. Cette étude ne met pas en évidence d'inflexion du niveau de ségrégation des groupes de revenus entre quartiers. Nous avons cherché à bâtir une base harmonisée de revenus localisés sur plus longue période, afin d'identifier d'éventuels effets des politiques publiques sur le niveau de ségrégation. En particulier, afin d'évaluer un effet causal du Programme national de rénovation urbaine sur la composition sociale des quartiers, il est nécessaire de disposer de données commençant avant le lancement des travaux de rénovation en 2003⁵.

Il existe aujourd'hui deux grandes sources pour mesurer les revenus à une échelle géographique fine. Ces deux sources sont basées sur les déclarations d'impôt des particuliers (principalement l'impôt sur le revenu, taxe d'habitation).

L'INSEE a constitué une base de données de revenus localisés « au carreau » de 2004 à 2020. Cette base qui s'appuie sur le dispositif RFL (revenus fiscaux localisés) pour les années 2004-2011 et sur le dispositif FILOSOFI (fichier localisé social et fiscal) pour les années ultérieures. La base ne peut remonter avant 2004, les premiers millésimes RFL n'étant pas localisables à un niveau géographique fin. L'intérêt de cette source est de découper le territoire en carreaux stables dans le temps et dont la taille peut être modulée (200 mètres de côté, un kilomètre de côté, etc).

La DGFIP et le SDES ont développé un système alternatif de localisation des revenus, appelé FILOCOM (fichier des logements et des communes). FILOCOM existe depuis 1995. Les données de revenu sont localisées grâce aux données cadastrales (section cadastrale ; parcelle cadastrale depuis 2017).

La base harmonisée des revenus localisés constituée par France Stratégie s'appuie exclusivement sur les données FILOCOM.

⁴ Voir Cusset *et al.* (2021), « [La meilleure répartition des logements sociaux a-t-elle fait progresser la mixité sociale ?](#) », *Point de vue*, France Stratégie.

⁵ Voir rapport de Guyon N. (2016), [Étude des effets de la rénovation urbaine sur l'évolution du bâti et du peuplement dans les quartiers ciblés entre 2004 et 2013](#), rapport du LIEPP.

La constitution de la base harmonisée des revenus localisés a nécessité un travail d'harmonisation important des données FILOCOM, afin d'obtenir des quartiers aux contours stables dans le temps.

Le reste du document se déroule comme suit : la partie 2 décrit les informations présentes dans les différents millésimes de FILOCOM, la partie 3 présente la façon dont les quartiers ont été constitués à partir des sections cadastrales.

2. Description des informations présentes dans les bases FILOCOM

Les données FILOCOM sont disponibles pour toutes les années impaires entre 1995 et 2019, ainsi qu'en 2010. Elles couvrent la France métropolitaine jusqu'en 2013 et la France entière depuis 2015. Nous présentons ici de manière synthétique les principales informations présentes dans les bases.

L'unité de base des données est le logement (les locaux à usage d'habitation au sens de la taxe d'habitation). Tous les logements de France métropolitaine sont inclus en principe dans FILOCOM, qu'ils soient vacants ou occupés. Pour chaque logement, on dispose de quatre types d'informations :

- Les caractéristiques du logement : localisation, nombre de pièces, année de construction, valeur cadastrale locative.
- Les caractéristiques du ménage occupant : composition du ménage, revenu déclaré à l'impôt sur le revenu, ancienneté d'occupation du logement, statut d'occupation (propriétaire/locataire), type d'occupation (résidence principale/secondaire).
- Les caractéristiques du propriétaire : statut juridique (organisme HLM, personnes morales ou physiques), domiciliation.
- Les caractéristiques de la copropriété, le cas échéant : nombre de logements et nombre d'appartements, présence d'un ascenseur⁶.

Chaque millésime contient les informations au 1^{er} janvier de l'année n, sauf pour les revenus (année n-1).

La localisation du logement est réalisée à partir des données cadastrales. De 1995 à 2019, on dispose de l'identifiant de la section cadastrale du logement. Les sections cadastrales sont des regroupements de portions de terrain unique appelées parcelles cadastrales. Chaque commune définit elle-même le découpage en sections cadastrales de son territoire à des fins administratives et d'aménagement.

En principe, l'identifiant d'une section est stable dans le temps. Toutefois, les sections associées à un même identifiant dans différents millésimes peuvent voir leur périmètre évoluer. Le découpage communal en sections bouge au cours du temps, souvent de manière marginale – découpage d'une section ou fusion de deux sections – parfois pour

⁶ La liste complète des variables du dernier millésime de FILOCOM peut être consultée à l'adresse suivante : <https://www.casd.eu/source/fichier-des-logements-par-communes/?tab=1>

redessiner une partie significative du plan cadastral communal. Un travail d'harmonisation est donc nécessaire si l'on veut suivre des quartiers de même contours au cours du temps, nous y reviendrons dans la partie 3.

Dans la plupart des millésimes de FILOCOM, pour des raisons de secret statistique, les logements localisés dans une section de moins de 11 logements n'ont pas d'identifiant de section et sont donc localisables au niveau communal uniquement.

Au cours du temps, la base FILOCOM a fait l'objet d'améliorations. Trois au moins affectent significativement les choix méthodologiques retenus dans l'étude :

Amélioration #1 – L'information sur la composition du ménage occupant chaque logement s'est améliorée au cours du temps. Ceci a des conséquences directes sur la notion de revenu observable dans les données. En 1995, on ne connaît pas le nombre de personnes du ménage, ni leur âge. En 1997, on connaît le nombre de personnes, mais pas leur âge. Depuis 1999, on dispose de ces deux informations.

Il est nécessaire de disposer du nombre et de l'âge des membres du ménage pour calculer des revenus ayant un sens d'un point de vue économique (revenu par adulte, revenu par unité de consommation par exemple), et permettant de répartir les individus par groupe de revenu. Cette opération ne peut être réalisée de manière satisfaisante que sur la période 1999-2019. Les données des millésimes 1995 et 1997 peuvent éventuellement être mobilisées pour calculer des évolutions du revenu global par quartier sur la période 1995-2019.

Amélioration #2 – À partir de 1999, tous les logements disposent d'un identifiant unique national. Cet identifiant permet de suivre un logement d'un millésime à l'autre. Il permet également de repérer les destructions et les constructions ayant lieu entre deux millésimes.

Comme l'a montré Nina Guyon (2016), on peut utiliser judicieusement l'identifiant unique national des logements pour suivre les découpages et les fusions de sections cadastrales, et aboutir à des sections cadastrales de périmètre constant au cours du temps.

Amélioration #3 – À partir de 2017, tous les logements sont associés à un code de section cadastrale, y compris si la section fait moins de 11 logements. On dispose également du code de parcelle cadastrale de chaque logement.

3. Méthode de constitution de sections cadastrales aux contours stables dans le temps

L'harmonisation des sections cadastrales est réalisée sur les 234 agglomérations de France métropolitaine appelées ci-après agglomérations de la politique de la ville, c'est-à-dire incluant au moins une zone urbaine sensible (ZUS) ou un quartier défini à l'article 6 de la loi de 2003 d'orientation et de programmation de la rénovation urbaine (ci-après quartiers article 6)⁷. Notre base harmonisée des revenus localisés est construite à partir des fichiers FILOCOM de 1999 à 2019. Les millésimes 1995 et 1997 n'ont pas été inclus dans notre

⁷ La liste complète des quartiers est téléchargeable sur le site data.gouv.fr.

base principale. D'une part, parce qu'il n'est pas possible de classer les individus de manière précise selon leur niveau de revenu pour ces deux années. Surtout, parce que les changements de périmètre des sections ne peuvent pas être observés avant l'apparition de l'identifiant de logement unique en 1999.

La méthode de constitution de sections cadastrales aux contours stables dans le temps, ci-après appelées « sections harmonisées », est tirée de Nina Guyon (2017)⁸. La méthode se base uniquement sur l'identifiant du logement qui est le même d'un millésime à l'autre à partir de 1999 ainsi que l'identifiant des sections cadastrales.

La partie a) décrit l'algorithme qui crée les sections cadastrales harmonisées, à contours stables entre 1999 et 2019.

La partie b) présente les retraitements effectués dans FILOCOM lorsque les identifiants de sections cadastrales associées aux logements sont incomplets ou non-renseignés.

La partie c) présente les résultats de l'exercice d'harmonisation obtenus sur les 234 agglomérations de la politique de la ville.

a) Harmonisation des identifiants de sections au cours du temps

Dans FILOCOM, les logements et les ménages fiscaux sont localisés au niveau de la section cadastrale. Au cours du temps, le nombre de sections cadastrales augmente tendanciellement. Au sein des 234 agglomérations de politique de la ville de France métropolitaine le nombre de sections cadastrales passe de 61 000 à 66 000 environ entre 1999 et 2019 (+9%). Le nombre de logements moyens par section augmente de 275 à 300 environ entre les deux dates. Les sections sont de « taille » (i.e. ont un nombre de logement) très hétérogènes : plus d'une sur cinq ont moins de 20 logements ; plus d'une sur cinq ont plus de 400 logements (voir Tableau 2). La variabilité du nombre de logements par section est d'ampleur comparable pour toutes les années d'observation.

Tableau 1 – Nombre de sections cadastrales et de logements dans FILOCOM au sein des agglomérations « politique de la ville »

Millésime	Nombre de sections cadastrales	Nombre de logements	Nombre moyen de logements par section
1999	60 815	16 644 053	274
2001	61 880	16 908 911	273
2003	62 558	17 151 930	274
2005	63 220	17 412 976	275
2007	63 536	17 709 357	279
2009	64 142	18 073 570	282
2010	64 357	18 258 184	284

⁸ Guyon N. (2017), « Quels effets de la rénovation urbaine sur les quartiers ciblés ? » *Policy Brief*, n° 29, février, LIEPP - Laboratoire interdisciplinaire d'évaluation des politiques publiques (Sciences Po).

Millésime	Nombre de sections cadastrales	Nombre de logements	Nombre moyen de logements par section
2011	64 587	18 434 091	285
2013	64 972	18 814 931	290
2015	65 435	19 169 358	293
2017	65 659	19 531 487	297
2019	66 138	19 916 836	301

Champ : 234 unités urbaines de France Métropolitaine (géographie 2020) comptant une zone urbaine sensible, ou un quartier « article 6 ».

Source : calculs France Stratégie à partir des données FILOCOM.

Tableau 2 – Distribution du nombre de logement par section cadastrale dans FILOCOM, en 1999 et 2019 au sein des agglomérations « politique de la ville »

	Nombre de logements par section cadastrale en 1999	Nombre de logements par section cadastrale en 2019
D1	5	6
D2	16	19
D3	33	42
D4	63	78
D5	109	131
D6	178	208
D7	279	314
D8	443	487
D9	778	842

Lecture : en 1999, 10 % des sections cadastrales comptaient plus de 778 logements.

Champ : 234 unités urbaines de France métropolitaine (géographie 2020) comptant une zone urbaine sensible, ou un quartier « article 6 ».

Source : calculs France Stratégie à partir des données FILOCOM.

La méthode d'harmonisation consiste à utiliser l'identifiant unique des logements observés sur des paires de deux millésimes consécutifs. On agrège toutes les sections qui « échangent » des logements entre elles au moins une fois entre 1999 et 2019. Par échange de logement, on entend le fait qu'un logement change de section cadastrale du fait d'un redécoupage géographique, qu'il s'agisse d'une scission de section ou d'une fusion partielle ou totale.

Notons qu'entre 1999 et 2019, dans les agglomérations « politique de la ville », 44 467 sections cadastrales n'échangent aucun logement avec une autre section sur deux années consécutives entre 1999 et 2019 ont un unique identifiant. Autrement dit, sur cette période, plus d'un quart des sections subissent des changements. Géographiquement, 38 % de la

surface des agglomérations considérées est constituée de sections dont le périmètre a changé au cours du temps.

Notre algorithme procède de manière itérative et chronologique :

- les sections qui échangent des logements entre 1999 et 2001 sont regroupées entre elles et se voient attribuer un identifiant de section commun,
- sur la base des échanges de logements entre les regroupements de sections ainsi obtenus et les autres sections cadastrales entre 2001 et 2003, on procède à un nouveau regroupement,
- on procède de la même manière pour toutes les années, jusqu'en 2019.

Ainsi l'algorithme aboutit au nombre minimal de regroupements avec des contours stables dans le temps. Il permet d'associer toute section « courante » de chaque année à une section « harmonisée », c'est-à-dire au contour géographique stable entre 1999 et 2019.

Les sections sont regroupées dans une section commune lorsqu'elles échangent des logements et que ces logements représentent 1 % de la section de départ ou 1% de la section d'arrivée. Ce principe, retenu pour le rapport du LIEPP en 2015, évite de regrouper entre-elles des sections marginalement affectées par le redécoupage du cadastre.

Encadré – Une limite importante de l'algorithme basé sur l'identifiant des logements

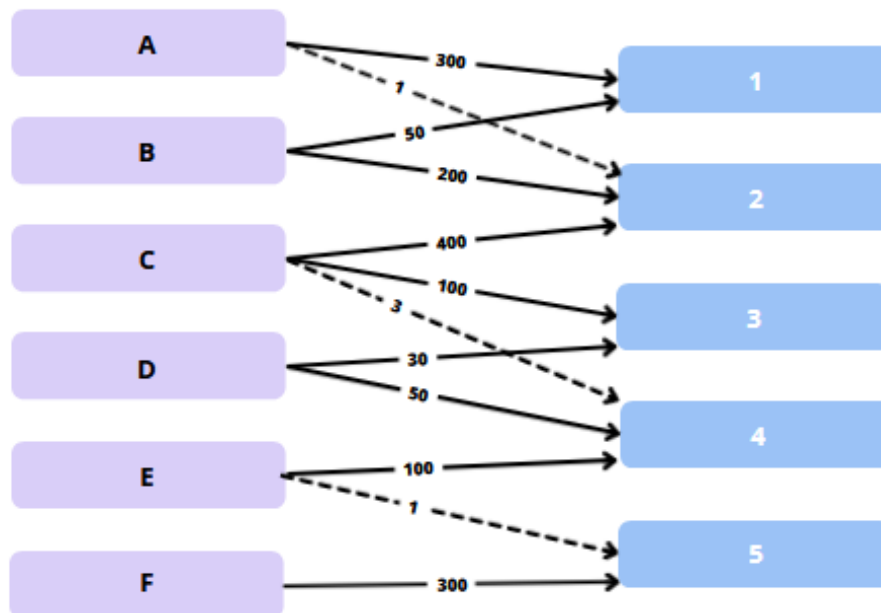
Certaines sections spécifiques ne peuvent pas être reliées à une section dans un autre millésime et sont donc exclues au terme de l'harmonisation. En effet, une contrainte importante de l'algorithme tient à la nécessité de disposer de logements existant dans deux millésimes consécutifs. En cas de destruction de la totalité des immeubles d'une section, ou de constructions d'immeuble sur une section auparavant vide, il est impossible d'associer une section à une autre dans le millésime postérieur, ou antérieur. Un autre cas de figure est celui des sections de moins de 11 logements, qui n'ont pas d'identifiants et sont également exclues par l'algorithme (même si ce problème a été en partie résolu, voir discussion autour du Tableau 4). Entre deux millésimes consécutifs, le nombre de sections exclues par l'algorithme est très faible. Néanmoins, sur l'ensemble de 11 millésimes, la probabilité qu'une section soit supprimée par l'algorithme n'est pas négligeable, y compris des sections comptant de nombreux logements certaines années⁹. Il en résulte une perte significative de logements par millésime à l'échelle des 234 agglomérations de politique de la ville. Néanmoins sur les communes comportant une ZUS ou un article 6, le taux de perte est négligeable (voir section finale du document).

⁹ Par exemple, une section rassemblant des terrains non construits en 1999, où quinze logements ont été créés 2003, puis cinquante en 2010 et à nouveau cinquante en 2017, sera exclue par l'algorithme.

Le Schéma 1 présente la façon dont nous agrégeons les sections entre elles sur la base des échanges de logements. Prenons l'exemple d'un ensemble de six sections d'une commune observée en 1999 : ces sections sont identifiées par une lettre, de A à F. L'identifiant unique des logements permet de connaître l'identifiant des sections d'appartenance des logements en 2003, pour les logements qui existent toujours à cette date. On constate que tous les logements situés dans les sections A à F en 2001 sont localisés dans les sections 1 à 6 en 2003.

Entre 1999 et 2001, en raison de modifications géographiques, certaines sections sont fusionnées et/ou scindées. Ces modifications entraînent les échanges de logements suivants.

**Schéma 1 – Exemple fictif de liens
entre les sections du millésime 1999 et du millésime 2001**



Lecture : parmi les 301 logements localisés dans la section « A » du millésime 1999 de FILOCOM, 300 sont localisés dans la section « 1 » du millésime 2001 de FILOCOM.

Source : schéma des auteurs

Les étapes de l'harmonisation des identifiants de section cadastrale sont les suivantes. Tout d'abord, nous faisons la liste de toutes les associations directes entre sections, quel que soit le nombre de logements échangés. Par exemple, nous associons la section A avec les sections 1 et 2. Puis on retire les associations entre sections représentant moins de 1 % du nombre de logement de la section de départ et d'arrivée. À l'issue de cette étape, la section A de 1999 n'est plus associée qu'à la section 1 de 2001. De la même manière, la section B est associée aux sections 1 et 2 ; C aux sections 2, 3 et 4, D à 3 et 4, E à 4 et F à 5.

Tableau 3 – Description de la boucle d'association entre sections sur deux années consécutives de FILOCOM

Sections en 1999	Sections associées en 2001	Sections associées en 2001 avec plus de 1 % logements partagés	Sections associées après Étape 1	Sections associées après Étape 2	Sections associées après Étape 3
A	1,2	1	1,2	1,2,3,4	1,2,3,4
B	1,2	1,2	1,2,3,4	1,2,3,4	1,2,3,4
C	2,3,4	2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
D	3,4	3,4	2,3,4	1,2,3,4	1,2,3,4
E	4,5	4	2,3,4	1,2,3,4	1,2,3,4
F	5	5	5	5	5
Nombre de sections associées ajoutées à chaque étape à chaque étape			7	4	0

Note : les chiffres en rouge indiquent quelles les sections associées ont été ajoutées à chaque étape de la boucle.

Les dernières colonnes du Tableau 3 représentent le cœur du processus d'harmonisation des sections cadastrales entre deux millésimes. L'objectif est de regrouper les sections si elles ont au moins une fois été reliées entre elles. On procède à l'aide d'une boucle qui opère successivement de la première à la dernière ligne du tableau. La boucle commence par la première ligne correspondant à la section A. Cette section est associée à 1. Or 1 est également associé à 2 dans la deuxième ligne du tableau. On ajoute donc la section 2 aux sections associées à A. La boucle passe ensuite à la deuxième ligne, correspondant à la section B et qui est associée aux sections 1 et 2. La section 2 étant elle-même reliée à la section 3 et à la section 4 (sur la troisième ligne du tableau), on ajoute ces dernières aux sections associées à B. On répète cette première itération sur toutes les lignes du tableau. Sept connexions ont été ajoutées à l'issue de cette première étape (représentées en rouge dans la colonne 4 du Tableau 3).

Les étapes suivantes recommencent ce même processus en partant des associations établies lors des étapes précédentes. L'harmonisation s'arrête lorsque aucune nouvelle association n'est ajoutée. Dans notre exemple, c'est bien le cas lors de la troisième étape : le processus d'association des sections des deux millésimes consécutifs est achevé. Dans notre exemple, tous les logements sont rassemblés en deux sections en 1999 (sections « A, B, C, D, E » et « F ») et 2001 (sections « 1, 2, 3, 4 » et « 5 »), dont le périmètre est stable. On obtient ainsi une table de passage entre l'identifiant propre à chaque millésime et un identifiant de section agrégée, c'est-à-dire n'échangeant pas de logements entre deux millésimes.

On répète ensuite le même processus avec tous les millésimes suivants. On commence par remplacer les identifiants des sections de 2001 avec les identifiants agrégés entre 1999 et 2001. Dans notre exemple, cela revient à renommer toutes les sections de 2001 en

« A, B, C, D, E » ou en « F ». On recommence alors toutes les étapes décrites précédemment pour obtenir des sections agrégées sur les millésimes 2001 et 2003.

On procède de la même façon pour chaque paire de millésimes consécutifs suivants (2003-2005, 2005-2007, ..., 2017-2019), jusqu'à obtenir une liste finale d'association de sections en 2019. Les tables de passage entre identifiants permettent de relier chaque section courante avec une section cette la liste finale.

b) Retraitements effectués des identifiants de sections cadastrales incomplets ou non renseignés

Redressement des codes « commune absorbée »

L'identifiant unique d'une section cadastrale dans un millésime de FILOCOM est constitué de trois éléments :

- le code de la commune (variable « DEPCOM » dans FILOCOM),
- le code dit de « commune absorbée » (variable « COMABS »)
- le code communal de la section castrale (variable « SECCA »)

Dans la très grande majorité des cas, le code « commune » absorbée est vide. L'identifiant de la section consiste donc uniquement en un code « commune » et un code de section. Seules les sections des communes découpées en arrondissements ou ayant absorbé d'autres communes dans le passé sont dotées d'un « comabs ». Ce code permet de distinguer deux sections ayant le même code dans deux communes fusionnées ou deux arrondissements différents. Prenons trois exemples :

- 69389_009_AA est une section cadastrale du 9^e arrondissement de Lyon.
- 62249_624_AB est une section cadastrale la commune de Douai-Lens, qui a été absorbée. On identifie cette modification géographique par son code de commune absorbée non vide.
- 93066_000_0A est une section cadastrale la commune de Saint-Denis, elle n'a pas été absorbée par une autre et ne possède pas d'arrondissement. Son code de commune absorbée est par défaut « 000 ».

Nous avons réalisé un travail de reconstitution du code commune absorbée, qui a été non ou mal renseigné pour certaines communes en 1999 et 2001.

Dans les millésimes 2003 à 2019 de FILOCOM, le code commune absorbée de Marseille est composé de quatre caractères. En 2001, le comabs des sections de Marseille est tronqué à trois caractères. Les sections cadastrales portant le même code au sein d'un même arrondissement ne sont donc pas différenciables.

Afin de résoudre ce problème, nous avons utilisé les logements existant en 2001 et 2003. Pour chaque logement renseigné dans les deux années, on remplace le comabs tronqué de

2001 par le comabs observée en 2003. À l'issue de cette première étape, 99,26 % des logements localisés à Marseille en 2001 se voient attribuer un nouveau comabs.

Pour les logements existant en 2001 et pas en 2003, on procède de la manière suivante. On classe tous les logements de Marseille par identifiant de logement. Si le logement est encadré dans la liste par deux logements de même comabs, on lui attribue cette comabs. À l'issue de cette deuxième étape, 99,98 % des logements de Marseille ont un nouveau comabs. Les 67 logements en 2001 sans comabs à Marseille sont supprimés de la base.

En ce qui concerne l'année 1999, la variable comabs est absente sur toute la base. Il faut la recréer pour l'ensemble des communes concernées. Nous adoptons la même méthode que pour le comabs tronqué de Marseille en 2001. À l'issue des deux étapes, nous avons attribué un comabs à 98,5 % des logements de 1999 présents dans des communes avec comabs en 2001.

Enfin, pour faire des représentations cartographiques, nous avons modifié l'identifiant de section cadastrale pour la commune de Marseille. Le code commune absorbée « comabs » de ces sections dans FILOCOM est différent de celui utilisé dans les données cadastrales de data.gouv. Nous avons créé une table de passage entre la variable comabs de FILOCOM et celle de data.gouv. Cela a été possible grâce aux identifiants de parcelles cadastrales présents dans les millésimes 2017 et 2019. Chaque section est composée de parcelles cadastrales qui correspondent à des portions de terrain appartenant à un même propriétaire. On a effectué une jointure des deux bases de données en utilisant l'identifiant de parcelles comme clé de correspondance. Cette opération permet de voir la correspondance entre les codes de commune absorbée des deux bases de données et ainsi de le modifier pour FILOCOM (voir annexe).

Redressement des logements sans identifiant de section

Pour des raisons de secret statistique, jusqu'en 2015, les logements situés dans une section cadastrale de moins de 11 logements n'ont pas d'identifiant de section dans FILOCOM. Cette limitation des données a été levée à partir de 2017. À partir de cette date, sauf exception, tous les logements sont rattachés à une section ainsi qu'à une parcelle cadastrale.

Les logements sans section représentent entre 0,2 et 0,3 % du total entre 1999 et 2015, et 0,01% en 2017 et 2019 (voir Tableau 4).

Même si ces logements sont peu nombreux, ils fragilisent l'exercice d'harmonisation des sections au cours du temps.

Nous attribuons un identifiant de sections cadastrales aux logements qui n'en ont pas. Ainsi, pour chaque année de FILOCOM, si le logement n'a pas d'identifiant de section alors on lui attribue l'identifiant de l'année suivante, si celui-ci est renseigné. S'il n'est pas renseigné alors on lui attribue celui de l'année d'après, et ainsi de suite.

Tableau 4 – Nombre de logements non associés à un identifiant de section avant et après retraitement

Millésime	Proportion de logements sans identifiant de section avant retraitement	Proportion de logements sans identifiant de section après retraitement
1999	0,28 %	0,05 %
2001	0,26 %	0,04 %
2003	0,26 %	0,04 %
2005	0,25 %	0,04 %
2007	0,24 %	0,03 %
2009	0,24 %	0,03 %
2010	0,24 %	0,03 %
2011	0,23 %	0,02 %
2013	0,23 %	0,02 %
2015	0,22 %	0,01 %
2017	0,01 %	0,01 %
2019	0,01 %	0,01 %

Champ : 234 unités urbaines de France métropolitaine (géographie 2020) comptant une zone urbaine sensible, ou un quartier « article 6 ».

Source : calcul des auteurs à partir des fichiers FILOCOM.

c) Résultat de l'algorithme d'harmonisation

À l'échelle des 234 agglomérations de la politique de la ville, la proportion de logements qui sont perdus au cours du processus d'harmonisation des sections cadastrales n'est pas négligeable : elle s'élève à environ 10 % du total de logements (voir Tableau 5). Cette proportion est relativement constante d'un millésime à l'autre.

Tableau 5 – Nombre de logements localisés dans une section harmonisée, au sein des agglomérations de la politique de la ville

Année	Nombre de logements total	Nombre de logements dans les sections cadastrales harmonisées	Proportion de logements manquants
1999	16 644 053	15 041 157	9,6 %
2001	16 908 911	15 281 934	9,6 %
2003	17 151 930	15 505 166	9,6 %
2005	17 412 976	15 738 435	9,6 %
2007	17 709 357	15 999 494	9,7 %
2009	18 073 570	16 321 293	9,7 %
2010	18 258 184	16 484 903	9,7 %
2011	18 434 091	16 641 891	9,7 %
2013	18 814 931	16 980 303	9,8 %

2015	19 169 358	17 302 967	9,7 %
2017	19 531 487	17 625 049	9,8 %
2019	19 916 836	17 974 234	9,8 %

Champ : 234 unités urbaines de France métropolitaine (géographie 2020) comptant une zone urbaine sensible, ou un quartier « article 6 ».

Source : calcul des auteurs à partir des fichiers FILOCOM.

Lorsqu'on retreint l'échantillon aux communes de la politique de la ville, le taux de perte chute fortement (voir Tableau 6). La proportion de logements manquants augmente au cours du temps, mais ne dépasse jamais 0,5 % du nombre de logements total de la base FILOCOM.

Tableau 6 – Nombre de logements localisés dans une section harmonisée, au sein des communes de la politique de la ville

Année	Nombre de logements total	Nombre de logements dans les sections cadastrales harmonisées	Proportion de logements manquants
1999	10 369 175	10 345 762	0,23 %
2001	10 497 338	10 476 596	0,20 %
2003	10 614 090	10 593 048	0,20 %
2005	10 736 835	10 716 149	0,19 %
2007	10 886 720	10 864 433	0,20 %
2009	11 073 885	11 048 947	0,23 %
2010	11 174 182	11 146 314	0,25 %
2011	11 266 737	11 236 432	0,27 %
2013	11 473 611	11 430 339	0,38 %
2015	11 659 126	11 612 430	0,40 %
2017	11 853 138	11 801 329	0,44 %
2019	12 061 720	12 004 966	0,47 %

Champ : 606 communes de France Métropolitaine (géographie 2020) comptant une zone urbaine sensible, ou un quartier « article 6 ».

Source : calcul des auteurs à partir des fichiers FILOCOM

La différence entre les deux échantillons tient à la densité du nombre de logement par section cadastrale. On rappelle (voir encadré plus haut) que l'algorithme exclut les sections ne contenant aucun logement ou très peu de logements dans au moins un millésime. Les zones peu denses à fort taux de destruction/création de logements sont particulièrement affectées, notamment dans le périurbain. À l'inverse, les communes de la politique de la ville étant constituées de quartiers denses et notamment de grands ensembles, la probabilité qu'une section ait très peu de logement à un moment donné est anecdotique.

On en déduit que l'algorithme d'harmonisation permet un suivi très satisfaisant des quartiers de la politique de la ville, mais qu'il induit une couverture imparfaite d'autres zones des agglomérations englobantes.

Annexe – Table de passage du code de commune absorbée de Marseille entre la base FILOCOM et le cadastre

Code commune	Code de commune absorbée FILOCOM	Code de commune absorbée data.gouv	Code commune	Code de commune absorbée FILOCOM	Code de commune absorbée data.gouv
13 201	101	801	13 210	1 003	857
13 201	102	802	13 210	1 004	858
13 201	103	803	13 210	1 005	859
13 201	104	804	13 210	1 006	860
13 201	105	805	13 211	1 101	861
13 201	106	806	13 211	1 102	862
13 202	201	807	13 211	1 103	863
13 202	202	808	13 211	1 104	864
13 202	203	809	13 211	1 105	865
13 202	204	810	13 211	1 106	866
13 203	301	811	13 211	1 107	867
13 203	302	812	13 211	1 108	868
13 203	303	813	13 211	1 109	869
13 203	304	814	13 211	1 110	870
13 204	401	815	13 211	1 111	871
13 204	402	816	13 212	1 201	872
13 204	403	817	13 212	1 202	873
13 204	404	818	13 212	1 203	874
13 205	501	819	13 212	1 204	875
13 205	502	820	13 212	1 205	876
13 205	503	821	13 212	1 206	877
13 205	504	822	13 212	1 207	878
13 206	601	823	13 213	1 301	879
13 206	602	824	13 213	1 302	880
13 206	603	825	13 213	1 303	881
13 206	604	826	13 213	1 304	882
13 206	605	827	13 213	1 305	883
13 206	606	828	13 213	1 306	884
13 207	701	829	13 213	1 307	885
13 207	702	830	13 213	1 308	886
13 207	703	831	13 213	1 309	887
13 207	704	832	13 213	1 310	888
13 207	705	833	13 213	1 311	889
13 207	706	834	13 214	1 401	890
13 207	707	835	13 214	1 402	891
13 208	801	836	13 214	1 403	892
13 208	802	837	13 214	1 404	893
13 208	803	838	13 214	1 405	894
13 208	804	839	13 214	1 406	895
13 208	805	840	13 214	1 407	896
13 208	806	841	13 215	1 501	897
13 208	807	842	13 215	1 502	898
13 208	808	843	13 215	1 503	899
13 208	809	844	13 215	1 504	900
13 208	810	845	13 215	1 505	901
13 209	901	846	13 215	1 506	902
13 209	902	847	13 215	1 507	903
13 209	903	848	13 215	1 508	904
13 209	904	849	13 215	1 509	905
13 209	905	850	13 215	1 510	906
13 209	906	851	13 215	1 511	907
13 209	907	852	13 216	1 601	908
13 209	908	853	13 216	1 602	909
13 209	909	854	13 216	1 603	910
13 210	1 001	855	13 216	1 604	911
13 210	1 002	856			

Source : table réalisée par les auteurs à partir du plan cadastral 2019 disponible sur data.gouv.