

FIDÉLI, L'INTÉGRATION DES SOURCES FISCALES DANS LES DONNÉES SOCIALES

Pierre Lamarche* et Stéfan Lollivier**

Les fichiers fiscaux se sont imposés dans la statistique publique comme des sources précieuses dont l'usage permet de satisfaire différents besoins du système d'information, à condition de faire l'objet d'un travail ambitieux de mise en cohérence. Le Fichier démographique sur les logements et les individus (Fidéli) produit par l'Insee à partir des données fiscales reflète bien ce potentiel associé à cette exigence ; en venant s'insérer en complément de sources telles que le recensement de la population, il permet d'éclairer des aspects des conditions de vie des individus jusqu'alors peu documentés, tant au niveau national qu'au niveau local. Ainsi, l'information finement localisée des individus permet de dresser un état des lieux plus complet, y compris au niveau infra-communal. Cette insertion se fait au prix d'un processus de production lourd et exigeant, et s'accompagne naturellement de contraintes d'utilisation assez élevées afin de préserver le secret statistique.

Pour rendre durable la présence de cette source dans le système d'information de la statistique publique, il faut dans un premier temps relever le défi posé par les réformes fiscales en cours, en particulier la suppression de la taxe d'habitation ; pour accroître son utilité, il faudra également poursuivre l'effort d'enrichissement, afin de tirer toujours plus parti de la profondeur de l'information fiscale. Enfin, Fidéli doit participer à terme à la mise en cohérence d'un ensemble encore plus vaste de sources administratives, et pas seulement fiscales.

 *Tax data have turned out as valuable sources for official statistics, as they fulfil a lot of needs of the information system, as long as they are produced through a demanding effort of data compilation. The Fichier Démographique des Logements et des Individus (Fidéli) produced by INSEE from fiscal data has proven the point. Usefully complementing traditional sources such as Census, it makes it possible to enlighten aspects of individuals' living conditions so far poorly documented, at national and at local level as well. Hence, the precisely geolocalized data makes it possible to renew the analysis also below municipality level. This completion is performed at the cost of a highly demanding and comprehensive production process while it also comes with strong dissemination conditions, so as to avoid unforeseen information disclosures.*

In order to make this statistical source sustainable, one should first address data issues posed by the ongoing fiscal reforms in France, especially the abolition of the Housing Tax. Second, one should enlarge its scope and encompass a always wider range of information coming from fiscal sources. Third and last, in the long run the source should fit in a more ambitious and comprehensive process involving larger parts of administrative – and not only fiscal – data.

* Chef de la division Logement, Insee,
pierre.lamarche@insee.fr

** Expert auprès du directeur général, Programme Répertoire des logements, Insee,
stefan.lollivier@insee.fr

L'administration fiscale, dans son activité de recouvrement de l'impôt, collecte à cette fin une grande quantité d'informations sur les individus et les logements. Ces données présentent un intérêt naturel pour le statisticien. Ainsi, l'usage des sources fiscales dans le système d'information des statistiques démographiques et sociales s'est imposé au cours de ces dernières décennies, avec en point d'orgue la production annuelle d'un ensemble de données permettant de répondre à des besoins à la fois très différents et très précis : le **Fichier démographique sur les logements et les individus (Fidéli)**.

La mise en cohérence des sources à l'origine de ces données illustre bien le processus qui permet de passer de données purement administratives à une information statistique de grande valeur pour le Service statistique public. Fidéli réalise en effet un traitement mutualisé des sources fiscales sur les logements et les personnes, et permet ainsi de mettre ces données à la disposition des chargés d'étude, des services en charge de l'échantillonnage des enquêtes, etc. Ces utilisateurs disposent chacun de « livrables » spécifiques (**figure 1 et encadré 1**) qui bénéficient de la même mise en cohérence des données administratives pour des finalités statistiques.

🌐 L'USAGE DES SOURCES ADMINISTRATIVES, DES PRATIQUES HÉTÉROGÈNES EN EUROPE

Si un observateur dressait un inventaire des différentes sources d'information utilisées par les instituts nationaux statistiques à travers le monde pour décrire les populations, il s'étonnerait de leur grande diversité. En Europe en particulier, il constaterait de très fortes différences d'un pays à l'autre.

“ Les pays nordiques (Danemark, Finlande, Suède, Norvège et Islande) construisent l'essentiel de leurs statistiques démographiques et sociales sur le recours intensif à des registres de population. ”

Les pays nordiques (Danemark, Finlande, Suède, Norvège et Islande) construisent l'essentiel de leurs statistiques démographiques et sociales sur le recours intensif à des registres de population, à tel point que le recensement de population se fait généralement sur cette base (Unece, 2007 ou Statistics Finland, 2004). Mais au-delà du dénombrement de population, les administrations de ces pays collectent et mettent à jour un spectre très large d'informations sur l'ensemble de la population, ce qui leur permet de produire

des analyses extrêmement variées. Une des clés de la réussite de ce modèle repose bien souvent dans l'adoption d'identifiants uniques pour repérer les unités d'intérêt – ici les individus et les logements – afin de faciliter les appariements entre les différentes sources. Très généralement, un tel système repose également sur deux piliers essentiels qui vont d'ailleurs assez naturellement de pair : un contexte législatif favorable, voire incitatif, ainsi qu'un grand degré d'acceptation et de confiance de la part de la population vis-à-vis d'un tel processus et des institutions qui sont les garantes de son bon usage.

De l'autre côté du spectre, de nombreux pays ont un recours encore assez limité aux registres et autres sources administratives, pour des raisons qui peuvent être tout autant d'ordre historique que d'ordre pratique. Ainsi, la constitution d'un registre, régulièrement actualisé, suppose la mise en place d'une infrastructure administrative pérenne et uniforme sur l'ensemble du territoire national ; un exemple très connu à ce sujet est l'absence de cadastre en Grèce.

De manière générale, les pays européens, en dehors de l'exemple nordique, ont pendant longtemps privilégié les techniques de sondage et les données d'enquêtes pour pouvoir éclairer les conditions de vie de la population ; les recensements ont été longtemps exclusivement basés sur de grandes opérations décennales de dénombrement de la population. De fait, la longue tradition de social-démocratie qui prédomine depuis longtemps dans les pays nordiques apporte un élément explicatif de l'usage intensif que font ces pays des registres pour éclairer les conditions sociales de leur population.

📍 EN FRANCE, UNE LENTE ÉVOLUTION MARQUÉE PAR LES IMPÉRATIFS DU RECENSEMENT

La France ne dispose pas de registre de population, et en matière de croisement de données d'origine administrative, l'administration et la statistique publique françaises ont longtemps été marquées par les polémiques dans les années 1970 autour du fichier Safari (Boucher, 1974)¹ ; cette affaire illustre bien la nécessité des deux piliers mentionnés plus haut comme pré-requis à un usage statistique des sources administratives, usage qui passe naturellement par l'appariement des différents registres. Le rejet par l'opinion publique d'un projet d'interconnexion de différents fichiers relatifs à la population a constitué par la suite un motif pour un recours assez exclusif à la collecte par enquête(s) : au fil du temps, l'Insee s'est constitué une expertise reconnue en la matière, permettant la description des grands phénomènes sociaux qui caractérisent la population française.

Les grandes enquêtes statistiques de l'Insee se sont historiquement adossées au recensement de population, qui permettait de constituer tous les dix ans environ une base de sondage exhaustive des logements, ménages et individus vivant en France. Les informations collectées à l'occasion de ces grandes opérations décennales étaient naturellement limitées, et la base de sondage était généralement complétée par des typologies socio-économiques des quartiers et communes de manière à accroître les possibilités de stratification. Ce modèle présentait alors un fort défaut d'actualisation, les échantillons sélectionnés alors en fin de cycle dans une base de données remontant à une dizaine d'années souffrant de problèmes liés entre autres aux mobilités résidentielles depuis la date du dernier recensement. La construction neuve était prise en compte en sélectionnant un échantillon complémentaire dans la base de logements neufs (Sitadel), mais sans connaissance des occupants de ces logements.

Le passage en 2004 à un recensement rotatif annuel a constitué de ce point de vue une amélioration substantielle puisqu'il a permis de constituer des bases de sondage renouvelées annuellement. Pour autant, l'usage du nouveau recensement de population comme base de sondage a également posé de nouveaux défis qu'il a fallu résoudre, comme le fait que le processus de recensement résultait lui-même d'un processus d'échantillonnage. Au-delà de ces questions méthodologiques, les unités primaires – c'est-à-dire les zones géographiques dans lesquelles sont sélectionnés aléatoirement les unités à enquêter – étaient menacées

1. Voir (Sénat, 2021) : « En 1974, le ministère de l'Intérieur avait bâti un fichier informatisé au nom évocateur : SAFARI, acronyme de système automatisé pour les fichiers administratifs et le répertoire des individus. Ce système prévoyait de créer une base de données centralisée de la population, en utilisant le fichier de sécurité sociale comme identifiant commun à tous les fichiers administratifs. Devant le tollé généralisé provoqué par ce projet – le journal *Le Monde* allant jusqu'à titrer « SAFARI, ou la chasse aux Français » –, le Premier ministre de l'époque [...] ne put que le retirer et créer dans la foulée une commission dite *Informatique et liberté*, chargée de proposer une réglementation sur l'utilisation des moyens informatiques ».

peu à peu d'épuisement ; la représentativité des échantillons sélectionnés s'en est trouvée petit à petit affectée². En outre, la rotation annuelle des échantillons en petites communes entraînait pour les enquêteurs des déplacements plus nombreux (Sillard *et alii*, 2020).

Parallèlement, la numérisation des données administratives, et en particulier fiscales, allait offrir de nouvelles opportunités. Dès 1996, l'Insee a pu disposer de la totalité des données fiscales relatives à la taxe d'habitation et à l'impôt sur le revenu. Une fois celles-ci

Encadré 1. Les livrables de Fidéli

Les produits issus de Fidéli sont multiples, et peuvent même faire l'objet d'une production ponctuelle lorsque le besoin s'en fait sentir. Les données produites de manière récurrente sont les suivantes :

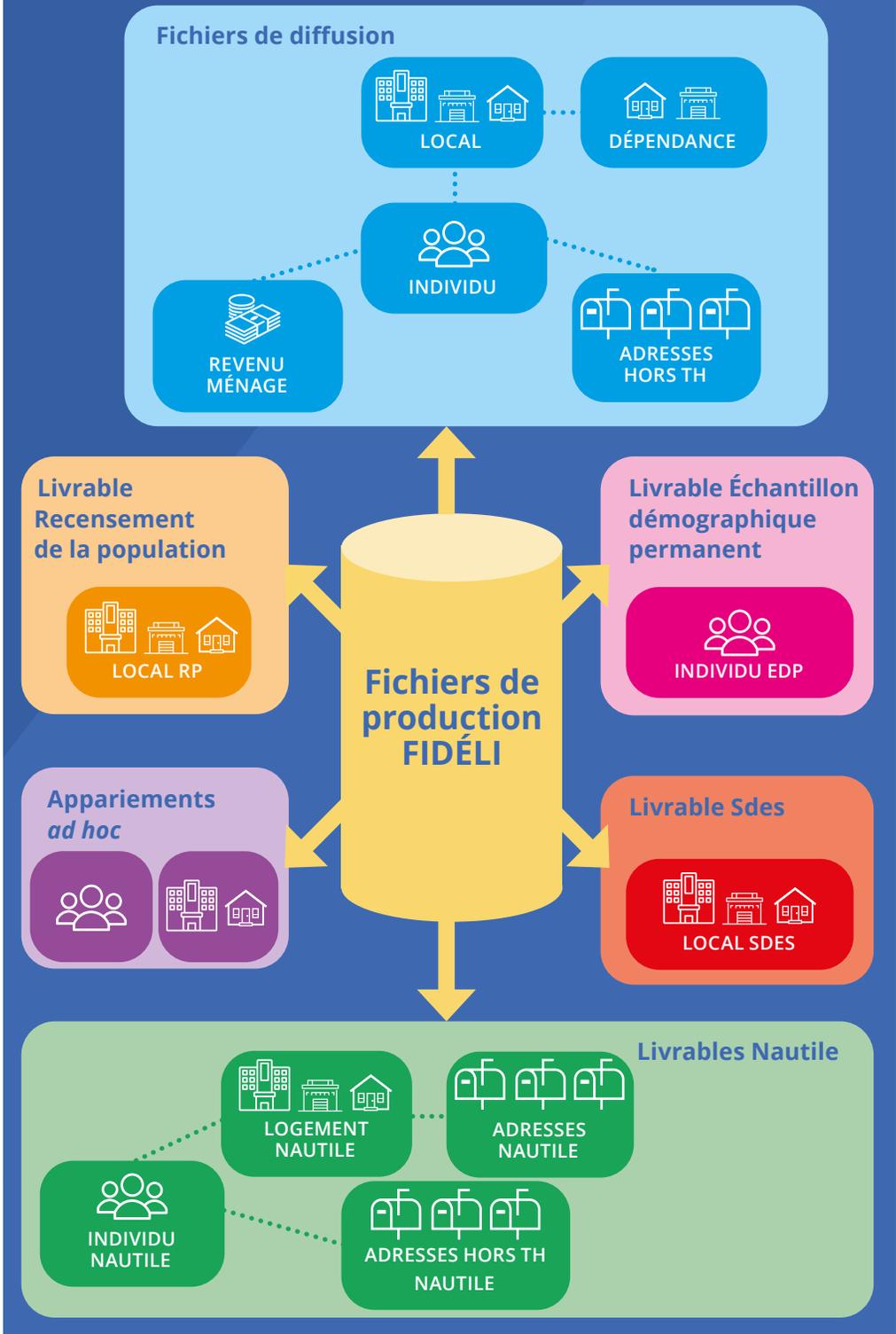
- **les fichiers de diffusion livrés au CASD***, qui sont les plus utilisés, sont constitués de cinq grandes tables, une table sur les locaux, une table sur les dépendances, une table sur les individus, une table sur les revenus des ménages localisés (par Filosofi) dans chaque logement recensé dans la table des locaux, ainsi qu'un complément, les adresses hors taxe d'habitation, qui correspond à l'ensemble des logements pour lesquels on retrouve des individus dans la source fiscale sans qu'ils ne soient connus à la taxe d'habitation. L'information contenue dans ces fichiers, très dense et précise, permet un grand nombre d'analyses, y compris au niveau local et constitue l'essentiel du socle sur lequel peuvent s'appuyer les utilisateurs de Fidéli. Comme il s'agit de fichiers de diffusion, l'information est anonymisée (au sens où l'ensemble des variables nominatives contenues dans les sources originelles sont supprimées), mais celle-ci demeure très largement identifiante, compte tenu de l'exhaustivité et de la précision de la source. Leur accès se fait dans un cadre très contraint, avec un examen des projets par le Comité du Secret, et une mise à disposition *via* le CASD pour les utilisateurs en dehors du Service statistique public, pour laquelle les analyses des données réalisées font l'objet d'un contrôle de confidentialité. Il s'agit des seuls fichiers mis à disposition en dehors du SSP ;
- **le livrable Nautile**, qui constitue la base de sondage des enquêtes auprès des ménages, contient l'information nécessaire pour la constitution des plans de sondage et l'établissement de fiches-adresses des individus ou logements échantillonnés. La seule finalité de ces fichiers est la constitution d'échantillons ; ils ne font pas l'objet de diffusion en dehors de la collecte de certaines enquêtes auprès des ménages, très encadrées par des conventions avec les services fiscaux (DGFIP) ;
- **un livrable spécifique pour le Sdes****, qui autorise un chaînage sur cinq ans (au lieu de deux dans le cas des fichiers de diffusion) des données sur les logements, de manière à reproduire une information similaire à celle produite par le fichier Filocom ;
- **un livrable permettant d'alimenter l'Échantillon démographique permanent ;**
- **un livrable utilisé par le recensement de population** pour anticiper la charge de la collecte dans les parties du territoire concernées par la prochaine Enquête annuelle de recensement (EAR).

* Centre d'accès sécurisé aux données (Gadouche, 2019).

** Services des données et études statistiques, service statistique des ministères chargés de l'environnement, de l'énergie, de la construction, du logement et des transports.

2. Ce d'autant que l'Insee a acquis une mission de service public en matière d'échantillonnage pour les enquêtes auprès des ménages : pour toute enquête ayant reçu le label d'intérêt statistique, l'Insee doit assurer le tirage de l'échantillon, généralement dans Nautile.

Figure 1. À chaque type d'utilisateur, son livrable



appariées avec l'enquête sur l'Emploi, l'Insee a pu annualiser la production de l'enquête sur les revenus fiscaux et répondre ainsi à une demande accrue d'informations sur les revenus des personnes et les inégalités.

L'utilisation à partir de 2009 des fichiers de la taxe d'habitation comme base de sondage de cette même enquête Emploi constituait alors la preuve que les sources fiscales peuvent représenter une alternative au recensement pour l'échantillonnage des enquêtes ménages. Le recours dans le même temps aux sources fiscales pour la description exhaustive, y compris au niveau local, de la distribution des revenus dans la population française³, et les outils de diffusion des résultats sous forme de données carroyées, ont renforcé la prise de conscience du potentiel des sources fiscales.

UN CONTEXTE DE PLUS EN PLUS PORTEUR

Dans le même temps, le contexte européen a vu l'émergence des sources administratives comme élément structurant du Système statistique européen (SSE), et particulièrement en France où les propriétés du recensement de population sous sa forme rotative pouvaient constituer un handicap (voir *supra* au sujet de la nature non-exhaustive du recensement rotatif). L'initiative GEOSTAT1 (Eurostat, 2018) a abouti ainsi à la diffusion de données de

« La loi pour une République numérique adoptée en 2016 a renforcé la possibilité de recours aux données sous forme numérique pour le besoin des enquêtes statistiques. »

population sur l'ensemble du territoire européen à l'échelle de carreaux de 1 kilomètre de côté. La volonté d'enrichir la diffusion de ces données carroyées avec de nouvelles variables, par exemple sur les revenus, consacre l'apport des sources administratives comme complément précieux aux sources traditionnelles de la statistique publique. De ce point de vue, le mouvement est global, tant en Europe, où des pays comme l'Espagne ou l'Allemagne (Bens et Schukraft, 2019) font également de plus en plus appel aux données administratives, mais également au Canada avec l'Environnement

de couplage de données sociales (Trainor et Trudeau, 2015), les Pays-Bas et le *System of social Statistical Datasets* (Bakker *et alii*, 2014), la Nouvelle-Zélande (Statistics New Zealand, 2014) ou encore l'Australie avec le *Multi-Agency Data Integration Project* (ABS, 2021).

Comme en Allemagne, le contexte législatif français accompagne le mouvement d'un recours plus massif aux sources administratives. La loi de 1951 sur l'Obligation, la coordination et le secret en matière de statistiques⁴ garantissait déjà de longue date l'accès pour le SSP aux données administratives. Mais la loi pour une République numérique adoptée en 2016 a renforcé la possibilité de recours aux données sous forme numérique pour le besoin des enquêtes statistiques. En outre, elle désigne l'Insee comme tiers de confiance⁵ dans le processus d'appariement entre différentes sources administratives. Le cadre législatif reste néanmoins assez contraignant, en particulier sur la question du principe de minimisation⁶, qui impose une durée de conservation des identifiants très courte, relativement aux besoins potentiels.

3. Dans le cadre de Revenus disponibles localisés (RDL) puis Filosofi (voir *infra*).

4. Voir les références juridiques en fin d'article.

5. Pour plus de précisions sur la notion de tiers de confiance, voir (Gadouche, 2019).

6. Il s'agit de l'obligation de ne disposer dans l'appariement que des données utiles à cet appariement ou au traitement pour lequel l'appariement doit donner lieu.

Dans la droite ligne de ces avancées et de cette prise de conscience du potentiel des données fiscales, le Service statistique public (SSP) s'est armé pour gagner en expertise et construire une information statistique de manière pérenne à partir de ces sources. Preuve de concept de l'intégration des sources fiscales dans le système d'information des statistiques démographiques et sociales, le **Fichier démographique sur les logements et les individus (Fidéli)** est ainsi apparu depuis 2016 dans ce système d'information : son objectif est de tirer parti des informations issues de l'administration fiscale sur l'impôt et les propriétés bâties pour compléter l'information déjà disponible sur le parc de logements et la démographie résidente. Au final, Fidéli constitue un bon exemple du travail de mise en cohérence, d'appariement et d'enrichissement de sources administratives permettant d'aboutir à une information statistique avec la constitution d'un objet dont la cohérence, l'exhaustivité et la variété d'informations disponibles sont essentielles pour son insertion dans le système d'information du SSP.

Le but de la construction du fichier est de disposer, à partir de plusieurs sources « brutes », d'une liste unique de logements d'habitation et d'une liste unique de personnes, puis de localiser ces personnes prioritairement dans leur logement principal, tout en regroupant les informations socio-démographiques les concernant.

DISPOSER D'UNE LISTE EXHAUSTIVE DES LOGEMENTS

Fidéli peut se définir comme **une base annuelle exhaustive de données statistiques sur les logements et de leurs occupants**, permettant d'éclairer le parc de logements ainsi que les mobilités résidentielles. Cette base de données repose tout d'abord sur les fichiers fiscaux sur le bâti, qu'il soit résidentiel ou non (**encadré 2**). Les données du bâti fournissent différents éléments :

- ① des éléments de repérage, comme l'adresse au cadastre (code Rivoli⁷, numéro de rue) ou encore l'adresse postale du propriétaire utilisée pour les correspondances par l'administration fiscale ;
- ① des informations sur la nature du propriétaire (personne physique ou morale) ;
- ① des informations sur la nature du bâti (maison, appartement, etc.), la superficie et le nombre de pièces, le nombre d'étages quand c'est pertinent, et d'autres éléments caractéristiques (présence d'un ascenseur, date de construction, etc.) ;
- ① des éléments de géolocalisation tels que la référence cadastrale⁸ ;
- ① et d'autres informations, telles que la présence de dépendances par exemple.

La fusion entre le fichier du bâti et les données de la taxe d'habitation permet d'enrichir la description des locaux, notamment sur le fait qu'ils sont occupés par le propriétaire ou un locataire, à titre de résidence principale, secondaire ou qu'il s'agit d'un logement vacant. La taxe d'habitation permet en outre de connaître les foyers fiscaux assujettis à la taxe pour un local donné, ce qui permet ultérieurement dans Fidéli d'en déduire les occupants. Il faut enfin être en mesure de caractériser, parmi les locaux, quels sont ceux qui peuvent être qualifiés de logements.

7. Avec Fantoir (fichier annuaire topographique initialisé réduit), anciennement fichier Rivoli (Répertoire informatisé des voies et lieux-dits), la Direction générale des finances publiques (DGFIP) recense, pour chaque commune, différents types de « voies » et leur attribue un identifiant appelé code Rivoli.

8. La référence cadastrale désigne une parcelle cadastrale de manière unique.

Fidéli introduit alors deux notions distinctes de logement au sens de la source fiscale :

- ① une première définition se base sur l'unique variable de nature du local, en considérant que le local d'habitation est nécessairement soit une maison, soit un appartement. Il s'agit de la définition centrale du logement dans Fidéli, et la plupart des traitements se rapportant au champ des logements réalisés dans Fidéli se font en cohérence avec cette définition ;
- ① une seconde définition plus proche du recensement, considère un univers plus large de natures de locaux (telles que des chambres de domestique, des pièces indépendantes, etc.), mais ne prend en compte que les informations issues de la taxe d'habitation pour définir le type d'occupation du logement.

Ces deux caractérisations des logements dans Fidéli coexistent, pour des usages distincts, l'un visant la cohérence interne des informations utilisées, l'autre cherchant à se rapprocher d'un concept similaire au recensement : lorsqu'on compare le volume de logements dans Fidéli selon la seconde définition avec celui mesuré dans le recensement, on obtient

Encadré 2. Les principales sources fiscales utilisées pour la constitution de Fidéli

Les sources fiscales mises en cohérence et intégrées dans le cadre de la production de Fidéli sont de différentes natures. Tout d'abord, il y a les sources sur le bâti :

- ① les données **Majic** (Mise à jour des informations cadastrales) fournissent toute l'information connue au cadastre, sur la nature du bâti, et sur la présence de dépendances. L'ensemble des locaux, y compris les logements, y sont recensés. Les données contiennent de l'information sur les caractéristiques des logements (nombre de pièces, superficie), ainsi que des immeubles dans le cadre de l'habitat collectif (nombre d'étages, nom d'usage éventuel). La source apporte également des informations sur la nature du propriétaire (particulier ou société), sur la localisation, telles que l'adresse au cadastre et l'adresse postale utilisée par l'administration fiscale pour les correspondances avec le propriétaire ;
- ① les données de la **taxe d'habitation** (fichier PLFC) sont partiellement redondantes avec celles de Majic, puisqu'elles décrivent les logements assujettis à la taxe d'habitation. Par ailleurs, elles contiennent une information sur le type d'occupation du logement (résidence principale, résidence secondaire ou occasionnelle, logement vacant), ainsi que la liste de l'ensemble des foyers fiscaux des occupants du logement (fournissant ainsi le lien logement occupant) ;
- ① le **FIP (fichier d'imposition des personnes)** quant à lui est une liste d'occurrences fiscales correspondant pour chacune d'entre elles à un ensemble d'individus connus par les services fiscaux, non nécessairement unique. Ce fichier permet d'associer un individu à un foyer fiscal ; il contient également de nombreuses informations identifiantes au sujet des individus (en particulier les traits d'identité, c'est-à-dire les noms, prénoms, date de naissance ou encore lieu de naissance). Ce fichier recense essentiellement les déclarants et leur éventuel conjoint, ainsi que les personnes adultes à charge ; une partie des individus mineurs manque, et est fournie par un autre fichier ;
- ① le **POTE (fichier permanent des occurrences de traitement des émissions)** est issu de la déclaration d'impôt sur le revenu (en particulier le formulaire 2042). Ce fichier complète les informations du FIP avec les statuts matrimoniaux des individus, et recense en théorie le nombre des personnes à charge des ménages. Ces données permettent également de fournir une information sur les revenus individualisables (salaires, pensions, indemnités chômage, etc.) pour chaque déclarant.

des différences de l'ordre de 1 % ; celles-ci s'expliquent en partie par l'absence dans les fichiers fiscaux des logements ordinaires qui ne sont pas soumis à la taxe d'habitation. Les écarts s'amplifient lorsqu'on s'intéresse aux résidences principales uniquement : là encore, cela est imputable au fait qu'un certain nombre de logements ordinaires ne sont pas assujettis à la taxe d'habitation, mais aussi à un certain retard dans l'enregistrement des logements récents dans les bases fiscales. Enfin, Fidéli comporte un nombre sensiblement plus élevé de logements vacants, car les sources fiscales tardent à prendre en compte les destructions de logements.

Une fois ces définitions établies, on poursuit la caractérisation des logements par **l'identification du parc des bailleurs sociaux, au moyen d'un appariement avec les données du Répertoire du parc locatif social (RPLS)**⁹. Cet appariement se déroule en plusieurs étapes, et implique un appariement sur le numéro Siren et sur la dénomination du bailleur. Une partie du parc est également repérée grâce aux informations d'exonérations provenant des fichiers du bâti.

Le parc de logements en France étant ainsi caractérisé, il faut ensuite lister l'ensemble des individus susceptibles de l'occuper et parvenir à relier chaque individu à un ou plusieurs logements.

DISPOSER D'UNE LISTE DES PERSONNES SANS DOUBLONS... —

Lorsque l'administration fiscale compile l'information dont elle dispose sur les individus, son objectif est le recouvrement de l'impôt, principalement de l'impôt sur le revenu et de la taxe d'habitation. En conséquence, si elle poursuit bien un objectif de complétude, elle ne se préoccupe pas de la question de la redondance ; de ce fait, un même individu peut être identifié par l'administration fiscale de manière non unique : il apparaît concrètement sur plusieurs lignes dans les fichiers fiscaux, sous des identifiants différents.

Le statisticien public poursuit, pour sa part, un double objectif d'exhaustivité et d'unicité dans le dénombrement : en clair, il s'agit de recenser précisément et en totalité les individus résidant sur le territoire français. Pour atteindre cet objectif, il faut repérer les doublons et les éliminer.

Les individus de 15 ans et plus vivant dans des foyers ayant déclaré des revenus ou payé la taxe d'habitation sont connus à partir du **fichier d'imposition des personnes (FIP)**. Celui-ci contient un grand nombre de variables décrivant les individus très précisément : date et lieu de naissance, sexe, nom (et éventuellement nom marital), prénoms, situation matrimoniale, et pour les individus décédés, année de décès. Comme il s'agit d'un fichier de gestion, plusieurs localisations peuvent figurer pour un même foyer fiscal (par exemple, ancienne et nouvelle adresse en cas de déménagement).

Les individus de moins de 15 ans ne sont connus que dans le **fichier relatif à l'impôt sur le revenu (POTE)**¹⁰, mais caractérisés seulement par leur année de naissance. L'information sur la situation matrimoniale des individus peut s'avérer incohérente entre les deux fichiers FIP et POTE. Fidéli applique alors des règles de mise en cohérence afin d'obtenir une donnée unique pour chaque individu connu de l'administration fiscale.

9. Pour plus d'information sur les données concernant le système d'information sur le logement, voir (Harmois et Lamarche, 2020).

10. « Fichier permanent des occurrences de traitement des émissions », élaboré par les services de la DGFIP à partir des émissions des avis d'imposition sur les revenus.

🕒 ... BÂTIR UN IDENTIFIANT NON SIGNIFIANT...

Pour les personnes de 15 ans et plus, les données fiscales contiennent un identifiant individuel. Mais celui-ci ne convient pas pour les finalités statistiques, et notamment du fait qu'environ un million de personnes sont présentes dans les données fiscales avec plusieurs identifiants fiscaux distincts. C'est la raison pour laquelle **Fidéli reconstruit un identifiant individuel non signifiant et spécifique à chaque année fiscale**, à partir de l'identifiant fiscal, mais aussi des éléments d'état-civil contenus dans FIP.

Cette opération nécessite de retravailler certaines informations : c'est notamment le cas du lieu de naissance, dont la recodification est l'un des éléments de standardisation de l'information fiscale les plus délicats. La variable est critique pour la constitution de l'information statistique, tant pour le repérage des doublons sur la base des traits d'identité que pour l'analyse *in fine*. Mais elle ne présente que peu d'intérêt pour l'administration fiscale, qui par conséquent n'apporte pas beaucoup d'attention à la qualité de sa collecte. Par ailleurs, l'information reste assez parcellaire selon le statut des individus vis-à-vis de l'administration fiscale. Ainsi, on dispose de peu de renseignements sur les personnes de moins de 15 ans à charge des foyers fiscaux. Cela pose, entre autres, un problème pour repérer les enfants en garde alternée. Autre difficulté : identifier puis localiser les étudiants, qui peuvent figurer dans le foyer fiscal de leurs parents en tant que personne à charge, et par ailleurs être connus de l'administration fiscale en tant que contribuables au titre de la taxe d'habitation.

🕒 ... CHAÎNER LES MILLÉSIMES...

Les mêmes informations relatives au repérage unique des individus dans les sources fiscales permettent également de retrouver la personne dans les fichiers fiscaux de l'année précédente, afin notamment d'appréhender les mobilités résidentielles et les changements d'état matrimoniaux. Le chaînage de l'information dans Fidéli entre deux années consécutives est fondamental, car il ouvre la porte à un ensemble d'analyses en matière de transition, que les données collectées dans le cadre du recensement de population ne permettaient jusqu'alors de traiter que de manière assez partielle. Ainsi, Fidéli permet d'observer les évolutions concomitantes en matière de revenu, de composition du ménage, d'éventuelles mobilités résidentielles (allant jusqu'à observer ces mobilités à l'aide de coordonnées géographiques).

🕒 ... ET RATTACHER LES INDIVIDUS À UN LOGEMENT (OU PLUSIEURS)

“ La résidence principale définit la localisation des individus. ”

Une fois qu'on dispose d'une liste cohérente d'individus et de logements, il convient de relier les individus à un ou plusieurs logements, de manière à les localiser, puis, grâce au suivi rétrospectif des individus, d'apprécier les mobilités résidentielles.

Pour faire un usage statistique de Fidéli, il est primordial de savoir où un individu habite, quelle est sa résidence principale : la résidence principale définit la localisation des individus. C'est à cette étape qu'intervient la connaissance des foyers fiscaux qui résident dans un même logement selon la taxe d'habitation.

Mais parce qu'ils peuvent appartenir à plusieurs foyers fiscaux liés à des adresses différentes, et parce que chaque foyer fiscal peut également être associé à plusieurs adresses, les individus ne sont pas naturellement liés à une seule adresse dans les sources fiscales. La procédure de localisation exige, ici encore, la mise en place de règles de décision permettant de lier un individu à une seule résidence principale. Fidéli utilise les informations fiscales pour déterminer la résidence principale, les règles liées aux contraintes fiscales étant parfois un peu différentes de celles adoptées par le recensement.

“ La procédure de localisation exige la mise en place de règles de décision permettant de lier un individu à une seule résidence principale. ”

Dans un premier temps, une étape de « nettoyage » est appliquée, afin de restreindre l'univers des possibles aux individus encore en vie à la date de référence, aux foyers connus à l'impôt sur le revenu et déclarant des revenus à l'adresse du logement, et aux foyers connus à la taxe d'habitation. Ensuite, si un individu appartient à plusieurs foyers fiscaux, on

applique des règles de priorisation entre les différents foyers connus par l'administration fiscale, afin de ne conserver qu'une seule localisation dans leur résidence principale, dès lors que celle-ci est connue¹¹. Lors de cette deuxième étape, il peut apparaître des individus figurant dans plusieurs foyers, dans l'un comme déclarant principal ou son conjoint, et dans les autres comme personne à charge (le cas typique est celui des jeunes adultes). La priorité est alors donnée en s'appuyant sur le statut de l'individu dans le foyer : il sera localisé dans le foyer où il est déclarant principal plutôt que dans celui où il est à charge (ses parents pour un jeune adulte).

Un foyer peut être localisé au sens de l'impôt sur le revenu à une adresse, et au sens de la taxe d'habitation à une autre adresse (**figure 2**). Si l'adresse de la taxe d'habitation est celle d'une résidence principale, c'est à cette adresse que l'individu est localisé. Si on ne lui connaît pas de résidence principale dans les sources fiscales, la personne est localisée à l'adresse de son imposition sur le revenu.

Par ailleurs, l'information sur l'adresse relative aux foyers fiscaux est multiple : dernière adresse connue, adresse de la résidence principale au sens de la taxe d'habitation, ou encore adresse utilisée par l'administration fiscale pour échanger avec les contribuables. Ces adresses seront plus ou moins pertinentes selon le but recherché au travers de la localisation des individus ; et elles présentent naturellement des imperfections en étant le résultat d'une gestion administrative, laquelle est de moins en moins fondée sur l'échange de courriers postaux.

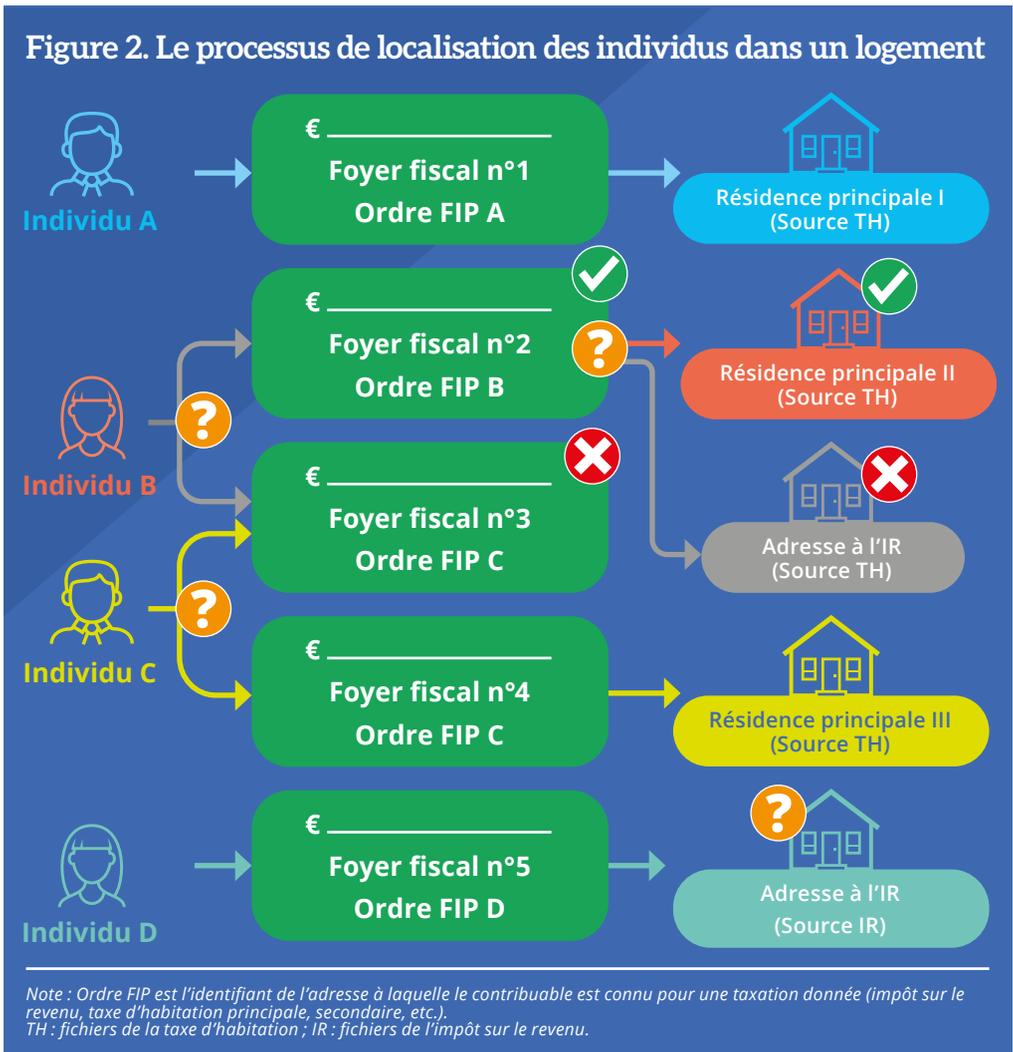
11. Si le foyer n'a pas de résidence principale au sens fiscal, il est localisé là où il déclare l'impôt sur le revenu.

REPERER LES MOBILITÉS RÉSIDENTIELLES...

Une fois la localisation des individus réalisée, la comparaison avec le millésime précédent permet de repérer les mobilités résidentielles au travers des changements d'adresse.

Les logements sont ensuite géolocalisés, sur la base d'abord des informations du cadastre, puis dans les cas problématiques, sur la base des éléments d'adressage (numéro, voie, etc.), avec une éventuelle procédure d'arbitrage dans les cas où les informations obtenues ne sont pas cohérentes. Fidéli construit ainsi une information nouvelle, sous forme de coordonnées géographiques, pour l'ensemble des logements du fichier. La géolocalisation permet également d'identifier les adresses appartenant à des IRIS¹² ou à des quartiers prioritaires de la politique de la ville. Cette information est extrêmement précieuse, car elle permet des analyses à un niveau local potentiellement très fin.

Figure 2. Le processus de localisation des individus dans un logement



12. Îlots regroupés pour l'information statistique, découpage infra-communal des communes de 5 000 habitants et plus.

Au final, le processus aboutit à la production d'une information globale et la plus cohérente possible permettant de lister l'ensemble des individus résidant sur le territoire national, en regard de leurs caractéristiques socio-démographiques et de leur localisation.

① UNE SOURCE COMPLÉMENTAIRE DANS LE SYSTÈME D'INFORMATION STATISTIQUE ACTUEL

Pour mieux cerner les logements ordinaires, les adresses dans Fidéli sont également enrichies avec les données du recensement sur les communautés ; de la même manière, les résidences hôtelières font également l'objet d'un traitement spécifique. Les données de Fidéli sont aussi rapprochées de Filosofi¹³, ce qui permet de disposer d'un revenu disponible et d'un niveau de vie pour les ménages vivant dans les logements pour lesquels l'information est disponible dans Filosofi.

“*Fidéli se positionne ainsi comme un complément extrêmement complet et riche du système d'information des statistiques démographiques et sociales.*”

L'ensemble de ces traitements est pensé de manière à se rapprocher le plus possible conceptuellement et quantitativement des informations traditionnellement collectées dans le cadre du recensement de population ou des enquêtes auprès des ménages. Fidéli se positionne ainsi comme un complément extrêmement complet et riche du système d'information des statistiques démographiques et sociales, répondant au besoin d'exhaustivité qu'induit la demande grandissante d'analyses fines au niveau local.

Fidéli est un produit intermédiaire, dans le sens où malgré sa grande richesse, il ne donne pas lieu à une diffusion propre en dehors des fichiers détails qui alimentent le Service statistique public ainsi que le monde académique. Par exemple, sur le dénombrement de la population à l'échelle d'une commune, Fidéli ne peut prétendre remplacer le recensement de population, qui fait foi ; comme dit précédemment, les sources fiscales ne sont pas construites sur un besoin originel de dénombrement de la population, mais de recouvrement de l'impôt. De ce point de vue, elles tendent à surestimer le nombre d'individus résidant sur le territoire national, tendance que le traitement et la mise en cohérence statistique réalisés dans le cadre de la production de Fidéli ne pallient malheureusement pas totalement. En revanche, elles permettent d'apporter des éclairages sur la population à l'échelle infra-communale, que le recensement sous sa forme moderne ne permet pas toujours, comme les mobilités résidentielles.

Fidéli, bien que contenant de nombreuses variables sur les revenus, ne remplace pas Filosofi : ce dernier reste la source de référence pour l'analyse de la distribution des revenus au niveau local ; en revanche, les variables de revenu disponibles dans Fidéli permettent d'envisager cette dimension comme un descripteur bien souvent pertinent pour les analyses rendues possible par le fichier.

13. Dispositif sur les revenus localisés sociaux et fiscaux. Les choix de localisation n'étant pas les mêmes pour Fidéli et Filosofi, l'appariement n'est que partiel.

📍 UNE RICHESSE EN MATIÈRE DE GÉORÉFÉRENCIEMENT ET DE VARIABLES DESCRIPTIVES

Depuis sa création, Fidéli a permis d'éclairer certains angles morts laissés par le recensement de population et les enquêtes ménages, notamment en ce qui concerne la géolocalisation dans les communes de moins de 10 000 habitants.

L'apport de Fidéli est illustré par un exemple récent : il s'agit de la possibilité de dénombrer de manière très précise, à l'aide des données de géolocalisation, le nombre d'individus vivant en zone à risque autour de l'implantation des centrales nucléaires.

Il est également possible d'évaluer de manière assez globale l'impact d'une submersion des zones littorales sur l'économie locale et la population, en utilisant les autres informations disponibles désormais au niveau local sur l'activité économique et l'emploi, et en les combinant avec les informations sur les caractéristiques socio-démographiques et les revenus des ménages vivant sur place (Brendler *et alii*, 2020).

Fidéli a aussi permis de savoir que les personnes habitant des quartiers prioritaires ont des mobilités résidentielles assez semblables aux autres résidents, contrairement à certaines idées reçues (Vicaire *et alii*, 2018).

Plus récemment, les mobilités observées au tout début de la crise sanitaire de 2020 ont pu être éclairées à l'aide des données de Fidéli, qui permettent de lier les individus non seulement à une résidence principale, mais également à d'éventuelles résidences secondaires et à leur localisation. Il a été ainsi possible d'expliquer les mouvements observés au travers des données de téléphonie mobile, à l'aune des informations que fournit Fidéli sur les résidences secondaires possédées par les ménages vivant en région parisienne.

In fine, l'information contenue dans Fidéli, sur les caractéristiques des logements et sur celles des ménages, que ce soit en termes de composition, de revenu ou de niveau de vie, combinée aux données sur les mobilités résidentielles d'une année sur l'autre, permet d'éclairer de manière assez complète les mobilités et de mieux en comprendre les ressorts.

📍 MAIS UN RISQUE RÉEL LIÉ À LA DIFFUSION DES DONNÉES LOCALISÉES

Toutefois, l'immense potentiel d'analyses que permet en théorie Fidéli masque de véritables contraintes dans l'usage qu'il est possible d'en faire. La richesse de ces données et le grand détail d'informations qu'elles contiennent révèlent en creux un risque accru de différenciation géographique du fait du grand nombre de zonages rendus possibles, et de la grande variété d'informations sur lesquels ces zonages peuvent porter.

Fidéli fait ainsi l'objet de conditions de diffusion très restrictives, dans la mesure où le spectre d'informations sur lequel il contient des variables est large. De ce fait, un usage mal contrôlé de ce fichier, par de multiples acteurs, souvent situés en dehors du Service statistique public et pour lesquels la coordination peut être plus difficile, pourrait amener au dévoilement involontaire d'informations individuelles sensibles. Son existence pose alors une exigence accrue pour le service producteur : exigence de vérification et exigence de centralisation des statistiques produites à partir des différents millésimes mis à disposition des utilisateurs. L'idée est donc de contraindre les utilisateurs *a priori* pour limiter le risque sur le secret statistique *a posteriori*.

Dans ce contexte, un autre facteur rend l'existence de Fidéli aisément valorisable par le monde de la recherche : la mise à disposition aux chercheurs des données détaillées dans le cadre d'une procédure et d'un accès sécurisés, grâce au Centre d'accès sécurisé aux données (Gadouche, 2019). En donnant l'assurance que les conditions de confidentialité et de diffusion sont strictement respectées, l'infrastructure du CASD permet un accès simplifié à des données telles que Fidéli, et démultiplie de ce fait les possibilités d'application dans le monde de la recherche.

📍 AU-DELÀ DES ÉTUDES, DE NOUVELLES MISSIONS POUR LA PRODUCTION STATISTIQUE

« *Fidéli constitue désormais la base de sondage de la plupart des enquêtes ménages menées dans le Service statistique public.* »

La finalité de Fidéli n'est pas uniquement de mettre à disposition des données pour les études. À l'issue du processus de traitement mutualisé des données fiscales, des livrables sont utilisés pour des finalités de production, par d'autres applications. C'est ainsi le cas pour l'échantillon démographique permanent¹⁴ ou quelques processus liés au recensement. Mais surtout, grâce à l'exhaustivité de la source, la qualité de son géoréférencement, et les très nombreuses variables descriptives des

individus, ménages et logements qu'elle contient, Fidéli constitue désormais la base de sondage de la plupart des enquêtes ménages menées dans le Service statistique public.

Originellement, cette mission était confiée au recensement de population ; le passage en 2004 à un recensement annuel rotatif a ouvert la voie d'un changement de paradigme : dans cette mutation, il s'est recentré sur le dénombrement de population en continu, en abandonnant de ce fait son caractère traditionnellement exhaustif qui est une propriété désirable pour une base de sondage (Sillard *et alii*, 2020).

Les sources fiscales se sont alors imposées comme un candidat naturel pour constituer une base de sondage, et ce d'autant plus qu'elles contiennent, du fait du recouvrement de la taxe d'habitation, un élément très intéressant pour la collecte des enquêtes auprès des ménages : le lien logement-occupant. Ce lien est central, car il permet l'identification précise sur le terrain de l'unité de collecte qui a été échantillonnée pour être enquêtée. Il est en effet difficile pour un enquêteur de repérer soit un logement sans en connaître l'identité de ses occupants, soit un individu ou un ménage sans en connaître le lieu d'habitation principal. Fidéli fournit une information de grande valeur pour les enquêtes ménages, en constituant une base de sondage mise à jour chaque année, et détaillant pour l'ensemble des individus résidant sur le territoire national leur localisation ainsi que de nombreuses variables contextuelles utiles à l'élaboration d'un plan de sondage pertinent. Fidéli présente ainsi différents avantages attendus pour une base de sondage de qualité : quasi-exhaustivité de l'information (y compris contextuelle), faibles défauts de couverture, mises à jour de l'information régulière, et pluralité des unités possiblement échantillonnées. En revanche, de nouveaux obstacles se font jour, comme la difficulté de repérage des logements dans le collectif, puisque les informations de rang de logement usuelles dans le recensement n'existent pas dans les sources fiscales.

14. Voir l'article d'Isabelle Robert-Bobée et Natacha Gualbert sur l'échantillon démographique permanent, dans ce même numéro.

Par ailleurs, le développement des collectes multi-modes pour ces enquêtes auprès des ménages nécessite de disposer, en amont de la collecte, de données de contact telles que des adresses électroniques et des numéros de téléphone : les sources fiscales en disposent et vont constituer une grande opportunité d'amélioration substantielle des conditions de collecte dans ce contexte¹⁵.

📍 LES PERSPECTIVES : D'ABORD MAINTENIR LES PROGRÈS RÉALISÉS..

Il n'est pas possible de parler de Fidéli sans évoquer la disparition des fichiers de la taxe d'habitation suite à la réforme fiscale entreprise en 2017, même si ces données sont moins utiles pour Fidéli, contrairement à Filosofi. Avec la suppression de la taxe d'habitation pour les résidences principales (Bur et Richard, 2018), on perd le lien entre occupant et logement, qui permet de reconstituer des ménages, mais on conserve tout le reste, qui a permis de réaliser la plupart des études évoquées plus haut.

Fidéli n'utilise le lien entre occupant et logement que de façon assez marginale, principalement pour reconstituer des ménages. En revanche, le lien entre occupant et logement est indispensable pour l'échantillonnage des enquêtes réalisés à partir de Fidéli. Compte tenu de l'importance de cet objectif, l'Insee a investi dans différents projets visant à assurer la pérennité de cette information sur différentes échelles de temps.

Dans un premier temps, un projet de court terme vise à effectuer une forme de ré-ingénierie de la chaîne de production du fichier Fidéli, de manière à rendre plus modulaire l'incorporation du lien logement-occupant et ainsi autoriser l'usage de sources alternatives contenant de l'information sur ce lien. À horizon de 2023, l'administration fiscale doit par ailleurs collecter l'information sur les occupants des logements car si la taxe d'habitation sur les résidences principales aura disparu, les taxes d'habitation sur les résidences secondaires ou la taxe sur les logements vacants existeront toujours par la suite ; il faudra donc bien être en mesure de distinguer les résidences principales et leurs occupants des autres logements. Cette nécessité de complétude du système d'information fiscale doit permettre d'assurer la pérennité de Fidéli dans les années à venir.

📍 ... ET POURSUIVRE L'INVESTISSEMENT SUR LES SOURCES FISCALES POUR ENRICHIR LE FICHIER

La force des données administratives, et en particulier fiscales, est qu'elles sont par essence porteuses d'autres types d'information que l'information collectée primitivement à but statistique, pourvu que l'on mette en œuvre le traitement approprié. Ainsi, les fichiers du bâti contiennent naturellement des informations sur l'identité du ou des propriétaires des logements ; il est alors possible d'enrichir les données existantes au travers du lien logement-propriétaire, ouvrant alors un nouvel horizon en matière d'analyses, cette fois sur la dimension patrimoniale.

15. Voir l'application dans le cadre de l'enquête Emploi dans l'article de François Guillaumat-Tailliet et Chloé Tavan dans ce même numéro.

En appariant les données du bâti avec les fichiers de Demandes de valeurs foncières¹⁶, il est même possible d'obtenir une valeur de marché pour l'ensemble des logements ayant fait l'objet d'une transaction dans les dernières années ; et à l'aide de modèles économétriques ou d'apprentissage correctement estimés, d'évaluer une valeur de marché pour l'ensemble des logements recensés dans Fidéli. Ces travaux ont déjà été entrepris, avec l'ambition d'incorporer dans les millésimes futurs les informations relatives au patrimoine immobilier des individus dans le champ de Fidéli. Les travaux relatifs au calcul du lien logement-proprétaire sont complexes, en particulier parce qu'ils nécessitent la mise en transparence des sociétés immobilières civiles, sans lesquelles l'information sur le patrimoine des ménages reste incomplète ; mais cette complexité est assez comparable à celle qui caractérise les traitements permettant la production de Fidéli sous sa forme actuelle, et peut donc être résolue avec des méthodes similaires. En revanche, ils renforcent l'utilité du fichier, et sa place de plus en plus centrale dans le système d'information des statistiques démographiques et sociales.

ET AU-DELÀ...

Au-delà, c'est une vision plus ambitieuse des sources administratives, et pas uniquement fiscales, qui prévaut afin d'assurer l'alimentation des besoins de la statistique démographique et sociale en matière de données localisées, variées et exhaustives. Il faut ainsi tendre vers une intégration beaucoup plus poussée des sources administratives pour assurer la résilience du système d'information vis-à-vis de la potentielle transformation, voire la disparition de certains fichiers. C'est à ce prix que les sources administratives occuperont toute la place qui peut être naturellement la leur dans le système statistique d'une économie administrée ; en contrepartie également, cette transformation suppose une vision beaucoup plus holistique, intégrée et cohérente du système d'information des statistiques démographiques et sociales.

16. DVF est un jeu de données sur les transactions immobilières en France produit par la Direction générale des finances publiques. Il est complémentaire des bases de données BIEN et PERVAL produites par les notaires. Voir (Harnois et Lamarche, 2020).

BIBLIOGRAPHIE

ABS, 2021. MADIP data and legislation. In : *site de l'Australian Bureau of Statistics*. [en ligne]. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP+data+and+legislation>.

BAKKER, Bart F. M., VAN ROOIJEN, Johan, VAN TOOR, Leo, 2014. The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. In : *Statistical Journal of the IAOS*. 2 avril 2014. Vol. 30, n° 4, pp. 411-424.

BENS, Arno et SCHUKRAFT Stefan, 2019. Modernisation des registres administratifs en Allemagne – Développements actuels et enjeux pour la statistique publique. In : *Courrier des statistiques*. [en ligne]. 27 juin 2019. Insee. N° N2, pp. 10-20. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4168390/courstat-2-3.pdf>.

BOUCHER, Philippe, 1974. Safari ou la chasse aux Français. In : *Le Monde*. [en ligne]. 21 mars 1974. Page 9. [Consulté le 25 mai 2021]. Disponible à l'adresse : https://www.lemonde.fr/archives/article/1974/03/21/une-division-de-l-informatique-est-creee-a-la-chancellerie-safari-ou-la-chasse-aux-francais_3086610_1819218.html.

BRENDLER, J., COMTE, S., LOUZA, T., MOUNCHIT, N., DARDAILLON, B., ROSE et V., PAILLETTE, É. 2020. *Plus de 100 000 résidents, logements et emplois concernés par le risque de submersion marine en Normandie*. [en ligne]. 7 décembre 2020. Insee Analyses Normandie, N°87. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/4989506>.

BUR, Dominique et RICHARD, Alain, 2018. *Mission Finances Locales. Rapport sur la refonte des finances locales*. [en ligne]. Mai 2018. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.vie-publique.fr/sites/default/files/rapport/pdf/184000278.pdf>.

EUROSTAT, 2018. Population grids. In : *Statistics explained*. [en ligne]. 16 juillet 2018. [Consulté le 25 mai 2021]. Disponible à l'adresse : https://ec.europa.eu/eurostat/statistics-explained/index.php/Population_grids.

GADOUCHE, Kamel, 2019. Le Centre d'accès sécurisé aux données (CASD), un service pour la data science et la recherche scientifique. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 76-92. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4254227/courstat-3-7.pdf>.

HARNOIS, Jérôme et LAMARCHE, Pierre, 2020. Le système statistique du logement – Étendue et perspectives. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee. N° N4, pp. 142-162. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497089/courstat-4-9.pdf>.

PADIEU, René, 2005. Grandes bases de données et protection des personnes. In : *Courrier des statistiques*. [en ligne]. Mars-juin 2005. Insee. N° 113-114, pp. 65-67. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.epsilon.insee.fr/jspui/bitstream/1/122548/1/cs113-114.pdf>.

SÉNAT, 2021. 1977 – 1978 : Le Sénat invente les autorités administratives indépendantes. In : *site du Sénat*. [en ligne]. Dossiers d'histoire. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.senat.fr/evenement/archives/D45/context.html>.

SILLARD, Patrick, FAIVRE, Sébastien, PALIOD, Nicolas et VINCENT, Ludovic, 2020. Pour les enquêtes auprès des ménages, l'Insee rénove ses échantillons. In : *Courrier des statistiques*. [en ligne]. 29 juin 2020. Insee. N° N4, pp. 81-100. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/fichier/4497081/courstat-4-6.pdf>.

STATISTICS FINLAND, 2004. *Use of registers and administrative data sources for statistical purposes. Best practices of Statistics Finland*. [en ligne]. [Consulté le 25 mai 2021]. Disponible à l'adresse : https://www.stat.fi/tup/julkaisut/kasikirjoja_45_en.pdf.

STATISTICS NEW ZEALAND, 2014. *Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project*. [en ligne]. Juin 2014. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/Linking-methodology-used-by-Statistics-New-Zealand-in-the-Integrated-Data-Infrastructure-project/linking-methodology-IDI-project.pdf>.

TRAINOR, Cathy et TRUDEAU, Richard, 2015. Environnement de couplage de données sociales (ECDS). In : *Conférence nationale du RCCDR des 5-6 novembre 2015*. [en ligne]. Statistique Canada et Réseau canadien des centres de données de recherche, atelier pré-conférence RCCDR. [Consulté le 25 mai 2021]. Disponible à l'adresse : https://crdcn.org/sites/default/files/4._sdle_overview_-_crdcn_pre-conference_workshop_nov_4_2015_french_final.pdf.

UNECE, 2007. *Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics*. [en ligne]. Nations Unies, New York et Genève. Séries Statistical standards and studies (Conference of European Statisticians). [Consulté le 25 mai 2021]. Disponible à l'adresse : https://digitallibrary.un.org/record/609979/files/Register_based_statistics_in_Nordic_countries.pdf.

VICAIRE, Vincent, SÉMÉCURBE, François, FAIVRE, Cynthia et DARRIAU, Valérie, 2018. Mobilité résidentielle entre 2015 et 2016 : un mouvement de même ampleur dans les quartiers prioritaires que dans le reste de la ville. In : *ONPV, Rapport annuel 2017*. [en ligne]. [Consulté le 25 mai 2021]. Disponible à l'adresse : http://www.onpv.fr/uploads/media_items/rapport-onpv-2017-synth%C3%A8se-stephanie-mas-1.original.pdf.

FONDEMENTS JURIDIQUES

Loi n° 51-711 du 7 juin 1951 sur l'Obligation, la coordination et le secret en matière de statistiques. In : *site de Légifrance*. [en ligne]. Mise à jour du 25 mars 2019. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573/2021-05-26/>.

Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. In : *site de Légifrance*. [en ligne]. Mise à jour du 9 décembre 2020. [Consulté le 25 mai 2021]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033202746/2020-12-09/>.