

Dossier suivi par :
LOONIS Vincent
Tél. : 01 41 17 68 48
Fax : 01 41 17 68 68
Mél : DG75-A050@insee.fr

Paris, le 9 novembre 2011
N° 4/DG75-A050/

Objet : Rénovation du questionnaire du recensement de la population présentation du projet de constitution d'un Répertoire Statistique des Logements à l'Insee.

1 Origine et contexte

Le système d'information disponible au moment des réflexions concernant la méthodologie du recensement rénové de la population a conduit à identifier l'adresse d'habitation comme niveau géographique le plus fin pouvant être repéré sans ambiguïté par un agent recenseur. Ce constat a façonné la structure de la base de sondage mobilisée pour la construction des échantillons des Enquêtes Annuelles de Recensement (EAR) : le Répertoire des Immeubles Localisés (RIL) que ce soit dans sa dimension unité statistique : l'adresse, ou dans la richesse des descripteurs associés à chaque unité, appelée traditionnellement information auxiliaire. Cette structure conditionne directement la méthodologie des EAR avec, par exemple, la nécessaire distinction entre « grandes adresses » et « petites adresses » afin de limiter les effets de grappe et améliorer la précision à budget fixé. La « qualité » de la base de sondages est ainsi un élément déterminant dans la mise en place des méthodes d'enquête, et par la même de la précision et du coût.

Depuis quelques années, dans le domaine connexe des enquêtes auprès des ménages, les données administratives, jusque là principalement mobilisées dans une optique de diffusion, ont montré des prédispositions à être également intégrées en amont du processus de production statistique. Depuis le 1^{er} trimestre 2009, l'échantillon de l'Enquête Emploi en Continu (EEC) issu originellement et traditionnellement du recensement, est progressivement remplacé par un échantillon issu du fichier de la Taxe d'Habitation (TH). Pour la première fois, au T4 2011, les ménages répondants à l'EEC ont tous été sélectionnés dans la source fiscale. Ce choix a été dicté en partie par des impératifs de délai au moment du débat social sur les statistiques du chômage de 2007 dont l'une des conséquences fut de procéder prestement à l'augmentation de 50% de la taille de l'échantillon de l'EEC. A ces impératifs de délais furent naturellement associés des impératifs de qualité statistique. Il est alors apparu que la source fiscale présentait globalement plus d'avantages que

les solutions « *traditionnelles* » sur l'ensemble des phases de la mise en place d'une enquête auprès des ménages, qu'elles soient statistiques : échantillonnage, correction de la non réponse, redressement, informatique : facilité de mise en place de la solution, ou de collecte : informations fournies aux enquêteurs pour le repérage des logements. Ces qualités ont par la suite conduit à retenir également la source fiscale pour l'échantillonnage de l'enquête Patrimoine, de l'Enquête PIAAC ou SIP.

Bien que mobilisées en amont pour des opérations importantes de l'Institut, le recours aux sources fiscales reste cependant ponctuel et non coordonné. Le programme « Répertoire des Logements » mis en place au 1^{er} août 2011 par l'Insee et piloté par Séfan Lollivier vise à autoriser cette mobilisation systématique des sources fiscales, mais plus généralement administratives, afin d'améliorer l'efficacité des enquêtes auprès des ménages dans leur ensemble et en particulier de la plus volumineuse d'entre elles : l'EAR. Le programme a vocation à abriter plusieurs projets. Le premier d'entre eux est la constitution concrète du répertoire, les autres ayant davantage pour objectif l'utilisation du répertoire.

2 Les avantages attendus et les principes de construction du répertoire

2.1 Les avantages attendus

L'expérience acquise lors des opérations telles que l'EEC, l'enquête patrimoine ou PIAAC ainsi que l'expertise méthodologique des fichiers fiscaux lors des travaux de préfiguration du répertoire menés de septembre 2010 à septembre 2011 à l'Insee permettent de mieux cerner les avantages à attendre du répertoire. Ces avantages s'articulent selon les qualités nécessaires d'une base de sondages à savoir :

L'exhaustivité : L'estimation du nombre de logements issue de la source fiscale est, en 2007, de 31,7 millions pour la France métropolitaine contre 31, 4 d'après le recensement : soit un écart de 0,8 %. Des fluctuations existent au niveau régional tout en restant modérées. Elles montrent cependant qu'un investissement est encore nécessaire dans le rapprochement des définitions fiscales et Insee du logement.

L'identifiabilité : les logements disposent d'un identifiant pérenne, et leur localisation est facilitée par des éléments d'adressage, patronymiques et surtout de référence cadastrale. Ce point constitue une innovation majeure à mettre en relation avec l'accès au Référentiel à Grande Échelle (RGE) de l'IGN. Ces références cadastrales assurent que même sans le nom des occupants et ou l'adresse du logement, il est possible de repérer un logement individuel grâce à son positionnement sur le fond de carte cadastrale.

La granularité désigne la capacité d'une base de sondage à autoriser l'échantillonnage d'unités statistiques dont la définition est aussi proche que possible de celle de l'unité que l'enquête cherche à observer. Par exemple, le RIL a une granularité trop importante par rapport à l'objectif des EAR qui est d'enquêter des logements et leurs occupants. La conséquence d'une granularité non adaptée est en générale une perte de précision. Les sources fiscales ont d'ors et déjà permis de



sélectionner des échantillons et de collecter de l'information auprès d'individus (PIAAC), des occupants de logements (Patrimoine) et d'occupants de tous les logements situés sur un même étage¹ (EEC). Des granularités plus fines que celle du RIL ont donc été validées empiriquement, même si le repérage des logements dans les immeubles collectifs devront faire l'objet d'une attention particulière.

La fraîcheur : la source fiscale relative au 1^{er} janvier de l'année N est disponible au début de l'année N+1. Elle contient cependant de l'information sur la localisation des ménages au moment de la déclaration de revenus c'est à dire à la fin du 2^{ème} trimestre N.

L'information auxiliaire est nécessaire pour mieux cibler les populations d'étude, optimiser les procédures d'échantillonnage, les méthodes de correction de la non réponse et celles de redressement des estimations. On notera que dans cette optique méthodologique, qui est celle du répertoire des logements, l'exigence de qualité est moindre que celle qui prévaudrait avec un objectif de diffusion. Il suffit que l'information soit corrélée au sujet d'étude pour qu'elle génère des gains statistiques. Une information ayant une définition différente de celle couramment retenue dans la sphère statistique ou présentant un taux d'erreur incompatible avec la diffusion peut, malgré tout, être génératrice de gains méthodologiques. Bien que l'objectif prioritaire du répertoire ne soit pas la diffusion, il est prévu de mesurer la qualité de l'information auxiliaire en procédant, par exemple, à des comparaisons macro et micro avec la collecte du recensement. Ce point est certainement central dans le cadre des réflexions du groupe de travail sur la rénovation du questionnaire du recensement.

Au final, la source fiscale présente une information auxiliaire riche est variée, en regard de celle actuellement disponible dans le RIL. Cette information est relative à au moins 2 niveaux :

- *les logements eux mêmes* : date d'achèvement, nombre de pièces, superficie, présence d'un ascenseur...
- *les occupants des logements* : état civil, statut matrimonial...

2.2 Les principes de construction

Le projet de constitution du répertoire des logements vise à permettre d'établir à **un rythme régulier**, au moins annuel, la liste des logements situés sur le territoire de la **France métropolitaine** et sur celui des **DOM** (hors Mayotte). Le champ des logements comprend les logements ordinaires, qu'ils soient principaux, secondaires ou vacants ainsi que **les logements en communautés**. A chaque logement sera associé une liste de descripteurs précisant :

- **sa localisation géographique** : adresse, étage, référence cadastrale...
- **ses caractéristiques** : sa nature (maison, appartement), son statut d'occupation (propriétaire, locataire...), sa catégorie...
- **les caractéristiques de ses occupants** : état-civil, statut matrimonial...

La liste est obtenue par **compilation automatique** de fichiers administratifs ou de sources internes à l'Insee. L'automatisme des traitements revêt plusieurs dimensions. Elle exclut tout recours direct à du travail de gestionnaire dans les Directions

¹ Dans le cas d'un logement individuel, l'étage est le logement lui-même.



Régionales de l'Insee. Elle confère ainsi au répertoire des logements, à la différence du RIL, **une qualité statistique**. C'est pourquoi le projet devrait être appelé **Répertoire Statistique des Logements (RSL)**. Le qualificatif de statistique a par ailleurs l'avantage de bien préciser que le RSL n'a pas vocation à être utilisé à des fins administratives, pour l'attribution d'un numéro Insee des logements par exemple.

La liste des fichiers susceptibles d'être mobilisés n'est pas définitive mais s'articulera autour d'une ossature fiscale à laquelle viendrait s'appuyer d'autres sources administratives ou internes à l'Insee. On rappelle ici que le projet est à un stade embryonnaire. Cela signifie que certaines sources doivent encore faire l'objet d'investigation et que le cadre juridique de l'ensemble doit être précisé.

- Comme pour le fichier FILOCOM, dont l'objectif est principalement orienté vers la diffusion, le fichier de la taxe d'habitation et celui des propriétés bâties devraient jouer un rôle majeur. Ils seraient enrichis par les informations relatives aux occupants contenues dans le Fichier d'Identification des Personnes (FIP) et dans une moindre mesure de l'Impôt sur le Revenu. Les déclarations en ligne pourraient quant à elles autoriser une mise à jour infra-annuelle..
- Dans le même esprit que les déclarations en ligne, le site de la Direction Générale de la Modernisation de l'Etat destiné à faciliter la prise en compte des déménagements par un ensemble d'administration est une orientation possible.
- Le Référentiel à Grande Échelle de l'IGN permettra d'obtenir via les références cadastrales les coordonnées X,Y des logements dans un premier temps, et l'établissement dans un second temps de documents de collecte améliorés.
- Dans un souci d'information, de concertation autour de Filocom des discussions sont prévues avec le Soes. Elles seront l'occasion d'étudier les opportunités liées aux permis de construire mais aussi au répertoire des logements locatifs des bailleurs sociaux.
- Afin de bénéficier des améliorations apportées par les gestionnaires du RIL en Direction Régionale, grâce notamment aux échanges avec les mairies, il est prévu d'intégrer dans la constitution du RSL au 1^{er} janvier de l'année N le RIL précédent millésimé au 1^{er} juillet de l'année N-1.
- Sous l'hypothèse d'une mise à jour infra-annuelle du RSL, l'intégration du RIL N peut aussi être envisagée.

Responsable du projet
de constitution du RSL

Vincent Loonis

